Title
Essays on liquidity of U.S. common stocks

Author
Nageswaran, Shalini

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays on Liquidity of U.S. Common Stocks

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Economics

by

Shalini Nageswaran

Committee in charge:

Professor Ross Starr, Chair
Professor Ivana Komunjer
Professor Bruce Lehmann
Professor Garey Ramey
Professor Allan Timmermann

2012

The Dissertation of Shalini Nageswaran is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2012

DEDICATION

For my family, my fiancé John, and my teachers.

EPIGRAPH

“Once you learn to quit, it becomes a habit.” – Vince Lombardi

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation was made possible due to helpful insights from my dissertation committee, friends, and seminar participants at UCSD and SEA. I especially thank John McAdams, my friend and colleague, for his continued support throughout the writing of this dissertation.

VITA

2006	Bachelor of Science (Honors), Purdue University, West Lafayette, Indiana
2010	Candidate of Philosophy, University of California, San Diego
2012	Doctor of Philosophy, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Essays on Liquidity of U.S. Common Stocks

by

Shalini Nageswaran

Doctor of Philosophy in Economics

University of California, San Diego, 2012

Professor Ross Starr, Chair

In Chapter 1, I find that stock characteristics do predict a stock's time-varying liquidity beta, i.e. its sensitivity to market, with the effect varying according to the assumed holding period using data on 30 small, medium, and large cap stocks between 1997 and 2002. I also find that liquidity is a priced factor for stock return even after controlling for market and stock measures of risk such as estimates of market volatility and stock level volatility. In order to mitigate problems arising from a small panel, I also test the returns model with ARCH errors on a larger sample of 2000 stocks.

Chapter 2 accounts for endogenous liquidity in a standard asset pricing model. Loss of liquidity, especially during times of crises, needs to be incorporated into models of financial assets so as to forecast returns correctly. To identify the effect of endogenous stock liquidity

on stock returns, an instrument is constructed from the NYSE, AMEX, and NASDAQ's decimalization program, which shrunk tick sizes from one-sixteenth of a dollar to one-hundredth of a dollar. Decimalization led to an increase in liquidity by allowing for narrower bid-ask spreads. Using daily price and quote data on U.S. common stocks, I find that as stocks becomes more illiquid, their future expected returns increase.

In Chapter 3, I propose two related measures for algorithmic trading constructed from the Disclosure of Order Execution Statistics data in order to study the effect of algorithmic trading on stock liquidity. The first is the average time taken to fill an order once it arrives in the market, known as fill time, and the second is the proportion of orders executed within ten seconds of order arrival at the NYSE. Since the decision to use algorithms in trading is an endogenous one, I use the NYSE's introduction of autoquotes in 2003 to identify the causal effect of algorithmic trading on a stock's liquidity. Using IV estimation, I find that a one second decrease in fill time narrows spreads by two basis points.

1 Pricing Asset Liquidity in U.S. Common Stocks

1.1 Introduction

Standard asset pricing theory states that securities with the same cash flows must have the same price. Frictionless markets with no transaction costs had been the norm in theories on asset pricing. However transaction costs can drive a wedge between gross returns and net returns. [1] In the presence of such costs, two identical stocks can have different prices without there being arbitrage opportunities due to different transaction costs. In this case, the security with the lower transaction cost is said to be more liquid. Investors holding such a security should demand returns lower than another security which has higher transaction costs and is less liquid. As Amihud et al (2005) note, on-the-run (newly issued) Treasuries trade at lower yields than almost identical off-the-run Treasuries. Also Amihud and Mendelson (1991) find that notes of the same cash flows trade at different prices.[2]

Finance literature historically has focused on a stock's correlation with the market return and market volatility to price an asset. As a result of this, we have a vast body of work related to the asset pricing literature. In the past couple of decades, increased availability of stock market data such as daily or intraday prices, bid-ask quotes, and other stock characteristics has enabled researchers to study how market microstructure induces differences in cross-sectional returns. Empirical studies can be divided into those that attribute expected returns to some systematic liquidity risk and others that attribute expected returns to a stock's level of liquidity. The systematic liquidity risk could be the stock's sensitivity to the aggregate

[1] Transaction costs here are a catchall for order processing costs, costs from information asymmetry, stock spread, and so on.
[2] Silber (1991) finds that shares that are restricted from trade for two years trade at an average discount of 30% relative to shares of the same company with identical dividends that can be traded freely. These are examples of some studies that have documented the evidence that liquidity in financial markets is an important aspect of asset pricing.

market liquidity the most recent measure of which was proposed by Pastor and Stambaugh (2003). Other papers include Chordia, Roll and Subrahmanyam (2000), Lo and Wang (2000), Hasbrouk and Seppi (2001), and Huberman and Halka (2002). In contrast, the stock's level of liquidity can be measured using stock-specific characteristics such as the bid-ask spread and the turnover rate. The literature tying level of liquidity to returns is based on the work of Amihud and Mendelson (1986), Brennan and Subrahmanyam (1996), Brennan, Chordia, and Subrahmanyam (1998) among others. For a complete survey of the theoretical and empirical literature on liquidity and asset prices, Amihud et al (2005) is an excellent resource.

This paper empirically answers the question of whether stock liquidity (level and movement with market liquidity) matters in explaining returns over and above what stock and market volatility can explain. A related question is whether the time-varying liquidity beta is predicted by stock-specific characteristics, with their effects varying over different holding periods. This brings together the two different literatures within the market microstructure literature. The main obstacle in pricing asset liquidity is that there is no one perfect measure of a stock's liquidity. Numerous measures have been proposed, most of them based on stock characteristics such as the bid-ask spread, volume traded, and number of shares outstanding. [3]

This paper uses the effective spread of a stock as the main liquidity variable. A stock's effective spread is defined as twice the absolute difference between log of stock's trade price and the log of mid-point of its bid-ask spread. This measure of liquidity has been considered more appropriate than the spread (the difference between the ask and the bid price) which may not convey much information with respect to the actual liquidity level of the stock when it is traded. The spread does not take in to account the sale price of the stock whereas effective spread does. The market wide liquidity measure used in this paper is the measure from Pastor

[3] Some papers that study measures of liquidity are Korajczyk and Sadka (2008) and Goyenko, Holden, and Trzcinka (2009).

and Stambaugh (2003), hereafter P&S (2003), whose monthly aggregate liquidity measure is a cross-sectional average of the individual stock liquidity proxy each month

Furthermore, many empirical studies have established that returns follow an auto-regressive and conditionally heteroskedastic (ARCH) process. In this paper the conditional variance (and volatility) is calculated through an ARCH process. The ARCH process was originally introduced by Engle (1982). However, since then many varieties of this model have been proposed. [4] Variations in time series dataset containing returns can be explained by many such models and fitting a wrong model will be disastrous when solving through maximum likelihood. The most appropriate of these ARCH models for financial data has been the class of asymmetric ARCH models. As of recently, these models include GJR-GARCH (Glosten, Jagannathan, and Runkle,1993), Exponential-ARCH (Nelson,1991), and the Power ARCH (Ding, Granger, and Engle, 1993). Asymmetric models are based on the reasoning that positive news about a stock's returns affects future returns differently than negative news does. The most simple of these models is the GJR model which uses only one parameter to estimate this asymmetric effect on a stock's conditional volatility.

A difficulty in the analysis of returns concerns the holding period. There is no definite data on how long stocks are held by investors. Barring institutional hurdles, one can hold a stock for any period between a day and the lifetime of the investor or the stock. Even with the advent of high-frequency data, most market microstructure studies only construct holding period returns over one particular time period, such as a day, month, or year. But there is good reason to believe that liquidity is priced differently based on the different holding periods. Amihud and Mendelson (1986) show that the liquidity effect on returns follows the shape of an inverted-U curve: it might not matter to very short term or very long term investors. But

[4] For a literature survey on the different ARCH models, see Engle (1995).

liquidity should be priced for investors with a medium range holding period of fewer than years and greater than days or weeks. With the use of high-frequency data, it is now possible to construct a more accurate (less noisy) measure of returns over different time periods using data sampled at intra-day frequencies. In this paper, daily, weekly, and monthly holding periods are constructed by averaging over periodical intra-day returns.[5] To summarize, the main questions being asked in this paper are as follows:

1. How does an individual stock's characteristics affect its comovement with the market liquidity factor? Specifically, what characteristics of stocks help predict its sensitivity to the market liquidity beta?

2. Are these effects significant over and above market and stock-level conditional variance?

We find that for weekly holding period returns, average dollar volume, after controlling for market liquidity, is consistently significant with an average coefficient of 2.3%. That is, a one standard deviation increase in average dollar volume of a stock with respect to the cross-section is associated with a 2.3% increase in returns. For monthly holding period, both average dollar volume and average number of shares outstanding are consistently significant. The coefficient on average volume in this case lies between 8.7 and 12.6% and that of number of shares outstanding lies between -7.4% and -17%. Average market variance, level of individual stock variance (ARCH-in-mean), and the Fama-French factors are used as control variables and are significant. The factors used for time-varying beta on market liquidity are all jointly significant (i.e. not equal to zero) in most of the specifications.

[5] Holding periods above a month are not considered because that would bring forth the necessity of including macroeconomic factors, and this paper strictly concerns with market microstructure. A project with holding periods over a month is being considered for future work.

These results imply that it is important to consider jointly the level of stock liquidity and its comovement with the market, after controlling for conditional variance in the market and stock level conditional variance. The empirical work done in this paper is closely related to that undertaken in Pastor and Stambaugh (2003), except that their aim is to obtain stock liquidity betas to use in tests of cross-sectional variation in returns. However, they do not control for changes in market volatility and this may be problematic. During times of economic distress, Watanabe and Watanabe (2008) find that stock's sensitivities to aggregate market liquidity increases and these "high liquidity beta" states, though short-lived, are characterized by heavy trade, high volatility, and wide dispersions in liquidity betas.

The rest of this paper is organized in the following way: Section II describes the data used, Section III specifies the models to run, Section IV analyses the regression results from the specified models, and tests other specifications. Section V concludes. The summary statistics and regression tables are attached in the appendix.

1.2 Data

1.2.1 Summary Statistics

The time period for the estimations is Jan. 1, 1997 to Dec. 31, 2002. The daily data on stock bid-ask quotes and final trade prices are obtained by averaging over hourly intra-day quotes and trade prices from the NYSE's Trade and Quotes (TAQ) database. [6]Stock characteristics such as cash dividends paid, volume traded, number of shares outstanding, and cumulative factor to adjust shares and prices (to account for share splits and new issues) are found in the Daily Stocks file from the Center for Research in Security Prices (CRSP)

[6] All the databases listed in this section can be accessed through Wharton Research Data Services (WRDS) from the Wharton School at University of Pennsylvania.

database. [7] The daily Fama-French factors and the Pastor-Stambaugh liquidity measure are obtained from Wharton Research Data Services (WRDS) as well.

Constructing returns from intra-day data allows us to avoid some of the pitfalls of using end-of-day prices. It is well known that closing prices behave erratically and using them to calculate returns may induce noise in the returns series. Thus, the availability of high frequency data enables us to get a better view of returns. Table 1 in the appendix shows the returns calculated using the intra-day price samples from WRDS and also the returns calculated using the daily holding period return available directly from CRPS. The daily holding period at time t by CRSP is calculated using the end of the day price at time t-1. Hence, the mean of the returns calculated from CRSP are significantly different from those calculated using intra-day averages. But the standard deviation of these returns using both method are very similar. The only discrepancy in the returns calculated using intra-day prices is that for two stocks, the daily return movement is very high (of a magnitude of 50-70%). A closer inspection of the data brings attention to the point that two of the stock paid magnanimous amount of dividends on those three days, and when we remove these three data points from the sample, there seems to be no difference in higher holding period returns. Regressions were carried out with and without these data points, and showed no difference in results. Table 1 also shows summary statistics on weekly and monthly returns.

Tables 2 and 3 summarize and correlate the averages of certain stock characteristics which will be used in the regression analysis (interacted with the marketwide liquidity factor). The standard deviations of these characteristics are high relative to the mean, which is an artifact of randomly choosing 30 stocks in each category of small, mid-cap, and large stocks.

---

[7] Adjustment for Price/Dividends is ((Price+Dividend)/(cfacpr)), where cfacpr is the cumulative factor to adjust prices. Adjustment for shares outstanding is ShrOutstan ding$*$cfacshr, where cfacshr is the cumulative factor to adjust shares.

In the correlation on Table 3, dollar volume and number of shares outstanding are highly correlated at 0.85. The liquidity measures effective spread and Gamma measure are weakly correlated at 0.27. And number of shares outstanding and effective spread are negatively correlated at -0.27. To see returns and spreads graphically, figure 1 displays the daily holding period return of the 30 stocks (not in any particular order). Returns are clustered over time, indicating ARCH effects. Returns for some stocks are more variable over time than others, as we have a mixture of small, mid, and large cap stocks.. And from figure 2, we can see that stocks with low return variability are also the stocks that are most likely to have very small spreads.

The NYSE and the NASDAQ underwent major changes with respect to tick sizes during the period from July 1997 to January 2000. To account for any changes in the mean of the returns during and following these events, dummies are included. The (1/8) tick is assumed as the base category. Hence, the dummy for (1/(16)) tick size switches on between July 24, 1997 through January 28, 2000. And the dummy for decimalization begins on January 29, 2000 through the end of the dataset in 2002.

1.2.2 Construction of Pastor-Stambaugh gamma measure for a stock

Pastor and Stambaugh (2003) calculate the liquidity cost of an individual stock in a given month with daily data on the stock's return and volume as follows:

$$R_{i,d+1,t}^{E} = R_{i,d+1,t} - R_{m,d+1,t}$$

$$R_{i,d+1,t}^{E} = \theta_{it} + \phi_{it} R_{i,d,t} + \gamma_{it} sign\left(R_{i,d,t}^{E}\right) . \nu_{i,d,t} + \epsilon_{i,d,t}, where\ d = 1, \ldots, 13$$

where $R_{i,d,t}$ is the stock's return on day d+1 in month t, $R_{m,d+1,t}$ is the CRSP value-weighted market return on day d+1 in month t; $\nu_{i,d,t}$ and is the dollar volume of the stock on day *d* in month *t*. This measure represents the average effect that a given stock's volume on day d has on the return on day *d+1*, when the volume is signed with the return on day *d*. Signed volume can be viewed roughly as order flow, in which case, if the stock's volume on a particular day leads to a price change in the opposite direction on the following day, then this implies that the stock is going through a low liquidity phase. Table 4 in the appendix displays the gamma measures calculated through this measure for individual stocks.

1.3 Model

1.3.1 Main Regression Model

The main regression takes on the following specification:

$$R_{it}^{E} = \beta_0 + \beta_{1it}L_t + \beta_2\left(\lambda.\hat{\sigma}_{Mkt,t}^2\right) + \beta_3\left(\lambda.\hat{\sigma}_{i,t}^2\right) + \beta_{4i}MKT_t + \beta_{5i}SMB_t + \beta_{6i}HML_t + \epsilon_{it}$$

$$\epsilon_{it} = e_{it} * \sigma_{it} \ where \ e_{it} \sim N(0,1) \ and \ Var(\epsilon_{it}) = \hat{\sigma}_{i,t}^2$$

$$\epsilon_{it} =_0 + \delta_1\epsilon_{it-1}^2 + \delta_2\epsilon_{it-1}^2\mathbf{1}_{\epsilon_{i,t-1}>0}$$

where R_{it}^{E} is the logarithmic return from holding the stock for a day from t-1 to t. The logarithmic returns are used so that we can additively construct higher holding period returns. L_t is the Pastor and Stambaugh (2003) market wide liquidity measure. MKT, SMB, and HML are the market excess return, return on small market capitalization portfolios minus the large market capitalization portfolios, and the return on value portfolios minus the growth portfolios. These are also called the three Fama French factors (Fama and French, 1993). Dummies for NYSE tick size changes are also included with the tick size of 1/8th as the base category. The error term is modeled with the asymmetric ARCH effect of Glosten, Jagannathan, and Runkle (1993). To simplify the estimation and as a good starting point, GJR(1)-ARCH(1) is used. This includes the ARCH(1) term, $\epsilon_{it-1}^2$and the asymmetric term that switches on when $\boldsymbol{\epsilon_{i,t-1} > 0}$. In the main regression, the conditional variance enters with the market price of risk term multiplied so as to individually identify the stock's sensitivity to this market variance. Note that restricting β_1, β_2, and β_3 to be zero and taking out the ARCH effects takes us back to the standard CAPM framework with the three Fama-French factors. The market conditional variance $\hat{\sigma}^2_{Mkt,t}$ is obtained through the regression where,

$$R_{CRSP}, t - R_{ft} = \alpha_0 + \lambda\, \sigma^2_{Mkt,t} + \epsilon_{CRSP,t}$$

where errors follow a GJR-ARCH(1) process, $R_{CRSP}, t - R_{ft}$ is the value-weighted return on all NYSE, AMEX, and NASDAQ common stocks minus the one-month Treasury bill rate. λ is the market price of risk and it is simply the coefficient of the regression of excess market return on the ARCH-in-mean term.

A stock's sensitivity to liquidity can be time varying. The time varying beta is modeled as follows:

$$\beta_{1,i,t-1} = \psi_{1,i} + \boldsymbol{\psi'_{2,i}}\mathbf{Z_{i,t-1}} + \epsilon_{3,t}$$

To account for this time varying beta coefficient, $\beta_1$in equation (1), some stock characteristics as denoted by the vector $\mathbf{Z_{i,t-1}}$ are used. [8] $\mathbf{Z_{i,t-1}}$ in equation (3) consists of the following: (1) the stock's historic liquidity beta, its sensitivity to the marketwide liquidity, calculate through an expanding regression with data until t-1, (2) the stock's average value of effective spread from days t-120 through t-1. Alternatively, the average value of γ_{it} was used from days t-120 through t-1, (3) the stock's average dollar volume (in millions of dollars) from days t-120 through t-1, (4) the stock's average return from days t-120 through t-1, (5) the stock's average price from t-120 through t-1and (6) the stock's average number of shares outstanding from t-

[8] These are the same stock characteristics used in Pastor and Stambaugh (2003).

120 through t-1. These series possess some explanatory appeal as the stock volume and level of liquidity can matter when we consider the liquidity risk of the stock. Controlling for number of shares outstanding and price would control for the size of the stock. And including average returns would account for any short run dynamics in the stock on the market. All stock characteristics are demeaned by subtracting the time series average through day t-1 of the cross-sectional average across stocks in each previous period. This scaling helps obtain the contribution of a stock characteristic to the liquidity beta while taking in to account the variation across stocks. The stock's average historic liquidity beta is obtained as β^L through the rolling regression of the daily excess return of stock i on the marketwide aggregate liquidity measure L, and the three Fama-French factors:

$$R_{it} - R_{ft} = \beta_0 + \beta_i^L L_t + \beta_i^{Mkt} MKT_t + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t$$

where the first 120 days of the sample is used to estimate initial coefficients and then expanded as each new day's information becomes available. To run the regression in (1), we can substitute equation (3) in to (1). This method of analyzing time-varying coefficients follows Shanken (1990), where we modify equation (1) to include six more independent variables that are constructed by interacting the stock level characteristics in (3) with the market liquidity factor L_t Regressions with holding period returns over a day, namely weekly holding period (5 trading days) and monthly holding period (approx. 20 trading days), are constructed by adding over the logarithmic daily excess returns for each stock from t through t+4 for monthly returns, and t through t+19 for monthly returns. That is,

$$Weekly\ Return_{it} = \sum_{s=t}^{s=t+4} Daily\ Return_{is}$$

$$Monthly\ Return_{it} = \sum_{s=t}^{s=t+19} Daily\ Return_{is}$$

1.3.2 Expected Coefficients

This subsection discusses the expected coefficient signs based on theory and past literature. Substituting for time-varying beta β_{1it} in Eqn.(1), we get:

$$R_{it}^{E} = \beta_0 + (\psi_{1,i} + \boldsymbol{\psi}'_{\mathbf{2,1}}\mathbf{Z}_{\mathbf{i,t-1}})L_t + \beta_2\left(\lambda.\,\hat{\sigma}_{Mkt,t}^{2}\right) + \beta_3\left(\lambda.\,\hat{\sigma}_{i,t}^{2}\right) + \beta_{4i}MKT_t + \beta_{5i}SMB_t + \beta_{6i}HML_t + \epsilon_{it}$$

The stock characteristics in the time-varying beta such as liquidity beta, effective spread, dollar volume, return, and shares outstanding are are six or one-month historical averages to account for any short-term effects of these variables. We expect the sign on the coefficient of liquidity beta to be positive if a stock with higher than average sensitivity to market liquidity commands a return premium over other stocks. Similarly, a stock with higher effective spread will be more illiquid, and hence will command a return premium. This effect translates as a positive coefficient on spread. The expected sign on the coefficient on volume is ambiguous as its sign depends on whether dollar volume proxies for a stock's liquidity, in which case it will have a negative coefficient. This is due to the fact that a stock with higher than average dollar volume is more liquid and hence will command less of a return premium. [9] But as noted in Campbell, Grossman, and Wang (1993), past levels of dollar volume can proxy for demand from liquidity or noise traders. Hence they find that stock price declines on a high-volume day is much more probable than a price increase and that these price decline days are associated with an increase in expected stock returns. The sign on the coefficient on Return will be positive if some sort of a mean reversion takes place. Size is usually calculated as stock price

[9] Studies that have found a negative effect of volume on returns include Amihud (2002), Chordia, Roll, and Subrahmanyam, and many others.

times the number of shares outstanding and originally regressions included both short term average of price and number of shares outstanding as stock characteristics. However, the average price did not contain any more information than the other included variables and hence was removed in all future regressions. The number of shares outstanding is a function of size and is expected to have a negative coefficient as found in previous literature on size effects. The sign on β_2 and β_3 should be positive, consistent with risk-return relationship. The coefficients on Fama-French factors must be positive and lie between 0 and 1. The tick size changes brought about during these years led to a decrease in the mean of returns due to an improvement of liquidity and we should see a negative coefficient on the tick size dummies if it were the case. Finally, since the ARCH model specified here deals with the asymmetric news effect, the sign on δ_2 is expected to be negative since negative news increases variance more than positive news. Or in this case, if the shock to a stock's return is positive, it should decrease the conditional variance of the stock.

1.4 Regression Results

Regression tables are reported in the appendix in Tables 5 through 10. Dependent variables are daily, weekly, and monthly excess returns, as specified in the table columns. Tables 5 reports the results from the pooled regression in equation (1) with the stock characteristics Z_{it-1} in equation (2) averaged over 6 months. Table 6 reports the results from a similar regression, except with the stock characteristics in equation (2) averaged over 1 month. Tables 7 through 10 report results from the unpooled regression in equation (1). Tables 7 and 8 report the results when the coefficients for all Fama-French factors and the market liquidity variable are firm-specific. Tables 9 and 10 reports the results when only the Fama-french coefficients are made firm-specific.

1.4.1 Discussion of Pooled Results

In the pooled regression in Table 5, we can compare the results across the holding periods. For the daily holding period return (not reported), none of the stock characteristics for the time varying liquidity beta are significant and have huge standard errors. Only the Fama-French factors are significant at the 1% level. Lack of significant results in this case is not very surprising as daily returns data almost always suffer from non-normality and non-stationarity. [10] This affects the sampling distribution of the coefficients. Indeed, an Augmented Dickey-Fuller test for panel data, as developed by Maddala and Wu (1999), run with a trend and up to 2 lags does not reject the null of non-stationarity or presence of an unit root (p-value = 1.000). Hence, the daily returns data is non-stationary and linear regression models such as in (1) may not be sufficient to model them. Thus, further regressions do not include the daily level results. [11] Thus we now concentrate on weekly and monthly return holding period regressions. A mathematical issue to be noted is that since the stock characteristics are demeaned as described in the previous section, the stock's coefficients for time-varying beta should be multiplied by the time series average through day t-1 of the cross-sectional standard deviation of that characteristic in each previous period. The reported tables reflect the coefficients multiplied by this factor. Thus, coefficients on the stock characteristics explaining time-varying liquidity beta is interpreted as the increase/decrease in future returns caused by a one-standard deviation increase/decrease in that characteristic of a stock with respect to the cross-section of stocks. The second column of Table 5 displays the weekly regression. A cross-sectional one standard deviation increase in average volume of a stock will increase returns by 2.22%. Similarly, a cross-sectional one standard deviation increase in average return of one stock in the past decreases future returns by 26.28%. While these numbers may seem large, given the variability of returns (as described in the Data section) coefficients of

[10] See Brown and Warner (1985).
[11] Alternate model specification to account for non-stationarity is being considered for future work.

such magnitude are plausible. The coefficient on spread is insignificant though positive. The Fama-French factors are significant and the betas on these variables fall between 0 and 1. The six-month average of market variance and the ARCH-in-mean variables are significant and positive. That is, a 1% increase in the historical average of the market variance increases returns by 6.7% and a similar increase in the stock's own variance increases its return by 0.002%.

In the monthly holding period column of Table 5, a one standard deviation increase in the average volume leads to a 12.6% higher return and the number of shares outstanding has a significantly negative coefficient of 17.11%. The expected coefficient on volume, as discussed in previous section, is ambiguous as volume can be a liquidity proxy or a proxy for demand of liquidity or noise traders. The positive coefficient on volume could be an implication of average past volume proxying for the demand of liquidity traders who would have to compensate informed traders through reduced prices, leading to higher future expected returns. As in the weekly case, the Fama-French factors, average market variance and the ARCH-in-mean terms are significant with expected signs.

Table 6 reports regression results based on the same regression run for Table 5 except that stock characteristics are averaged over 1 month instead of 6 months. The decision to run these set of regression arises from the fact that the six month average made sense in P&S (2003) as their dependent variable was constructed at a monthly rate. But in this paper, since we consider both monthly and weekly holding periods, a short term average of 1 month may be necessary instead of longer term average of six months. Table 6 displays only the weekly holding period regression. The monthly regression in this case did not achieve convergence. Furthermore, it turns out that the monthly averages do not really help in predicting liquidity betas. Only the Fama-French factors and the market and individual variances are significant.

In all of the above regression, dummies for tick size changes on the NYSE were included. But the decimalization dummy is dropped due to collinearity and the dummy size for tick size of 1/16 is not significant.

The pooled regressions are run following the methodology in P&S (2003). While the magnitudes are not directly comparable due to the usage of a smaller sample and different time period, the coefficient signs obtained here match with theirs. The signs for volume and share turnover match with their results whenever significant. And the negative sign on average returns matches with theirs, though their coefficient on average return is not significant across different time periods.

1.4.2 Discussion of Regressions with Firm-Specific Coefficients

P&S (2003) run pooled regression due to limitations on data and to increase precision. But given that we have more observations per stock and improved computing abilities, non-pooled regressions are run. Tables 7 through 10 display the same regression as before, except that we now have the coefficient on market liquidity β_{1,t} and Fama-French factors vary with the firm. In Tables 7 and 8 both market liquidity and Fama-French factor coefficients vary by firm, whereas in Tables 9 and 10, only the Fama-French factor coefficients vary by firm.

In Table 7, regression on weekly holding period return yield similar results as before. Dollar volume is significant and positive at the 5% level. The individual stock coefficients are not displayed as there are 120 of them. However, the Fama-French factors are significant for almost all of the stocks. Average conditional market variance is significant and positive, in accordance with the theory that increased market risk increases returns. The tick-size dummy for (1/(16)) is significant, though it has the opposite sign. A negative sign would be expected on tick-size dummies as these changes signified decreasing transaction costs and thus returns would be expected to fall.

Results for monthly returns are more promising with the average spread, dollar volume, and number of shares outstanding being significant. A one standard deviation increase in the average spread of a stock increases returns by 3.6% while the increase in dollar volume increases returns by 11.4% and that of number of shares outstanding decreases return by 7.4%. These variables are significant even after controlling for market and individual risk (variance).

Table 8, which has the stock characteristics averaged over one month instead of six months, presents similar results for weekly returns. Instead of dollar volume, number of shares outstanding is significant but positive. Tick sizes show the right sign of negative and only the decimalization dummy is significant at the 10% level. Results for monthly returns are significant and most variables have expected signs except for the average spread. An examination of standard errors shows that the decimalization dummy has missing standard errors and this might be a problem econometrically. This might arise due to misspecification or problems with maximum likelihood convergence due to collinearity or reduced rank. Or it is entirely possible that tick size dummies are accounting for some other mean return behavior (some sort of time trend) other than what they are intended for.

Tables 9 and 10 constrain the market liquidity coefficient too be the same across all stocks. This is akin to removing the firm fixed effect in the equation for time-varying liquidity beta β_{1t}, thus reducing the number of parameters by 30. The results are similar to the case of pooled regressions. In weekly returns, average volume, returns, and number of shares outstanding are significant and have the expected signs of positive, negative, and negative, respectively. The magnitude are very similar to that of Table 5. Monthly returns show the historical average liquidity beta to positively predict future returns. Specifically, a cross-sectional one standard deviation increase in the average historic liquidity beta predicts higher returns of about 5%. Average volume and number of shares outstanding are significant as before. Market and individual variances are significant, though individual variance is negatively correlated with returns, as opposed to being positive. Table 10 repeats the regression in the previous table, except with one-month averaging. The results are similar for weekly returns, and for monthly returns, we face the same problem of convergence. It is interesting to note that this mis-specification problem only changes the coefficient sign on average spread. Hence, it could be that one-month average of spread is highly correlated with standard deviation of stock returns obtained through ARCH.

1.4.3 General Discussion of Results and Discussion of Results from Alternate Dataset

To summarize the results so far, we look at variables that are consistently significant across specifications. When the dependent variable is weekly holding period returns, average dollar volume is a consistent predictor of returns, holding market liquidity constant, and has a positive coefficient in the range of 2.22-2.39% on returns. The average of the market variance times the market price of risk affects returns positively and is in the range 4.2-7.5%. When monthly holding period is the dependent variable, average dollar volume has a positive coefficient and a one standard-deviation increase increases returns anywhere in the range of 8.7-12.6%, while number of shares outstanding affects returns negatively by a factor of 7.4-17%, after controlling for market liquidity L_{t}. Average market variance also affected returns positively in the wide range between 7.5 to 32%. The Fama-French factors are also significant in all the regressions. The substantial difference in the liquidity beta and return predictors for weekly and monthly returns could arise due to the fact that number of shares outstanding for a stock substantively differs only by month and hence does not vary enough to be able to explain the variation in weekly returns. The magnitude of the coefficients are higher for the monthly case due to the fact that monthly returns have a higher mean and standard deviation than weekly holding period return.

On the other hand, some variables are not consistently significant. Effective spread is not significant across most of the specifications, save for three of the monthly holding period regressions in Tables 7, 8, and 10. These regression results also have a higher magnitude for the coefficient on market variance. Hence, the market variance could be influencing the coefficient on spread, even flipping the sign in one case. The insignificance of spread, though surprising at first, gives rise to the notion that liquidity effects arising from volume and number of shares outstanding override the effect of effective spread. Indeed, it is true that effective spread is not the only measure of liquidity and that many others have been proposed that are a function of volume and number of shares outstanding. Also, the regressions in this paper are run using data from a period when spreads are falling greatly and cannot explain variation in returns as well as they used to. This might be an important thing to remember for the future when researchers try to come up with new measures of liquidity.

Dummies for decimalization and $1/16^{th}$ tick size changes are expected to have negative coefficients if these tick size changes indeed increased liquidity and thus lowered returns. In the specification where one-month averages of stock characteristics are used for predicting liquidity beta, the decimalization dummy is significant, has a negative sign, and decreases returns by as much as 1.5% in weekly holding periods and 6% in monthly holding periods. However, the specifications using six-month averages of the stock-specific variables dropped the decimalization dummy due to collinearity and the (1/(16))^{th} tick size has the opposite sign in the cases where it is significant.

Finally it is imperative to test if the assumption of time-varying beta is valid and if aggregate liquidity is priced in the asset markets. Thus, we run joint F-tests testing the null hypothesis that the coefficients of stock-specific predictors of liquidity beta are zero. The null hypothesis that they are all jointly zero is roundly rejected at both 1% and 5% levels (not reported in paper) in non-pooled regressions in Tables 7 through 10 at both weekly and monthly holding period levels. In the pooled regression, the null hypothesis is rejected only for the specification with six-month average of stock characteristics. Hence, there is substantial evidence to conclude that the liquidity betas are time-varying and that most of the stock characteristics chosen to predict these betas do substantively predict them. Another round of F-tests are used to test if aggregate market liquidity as a whole is a priced variable. This is done by testing the null hypothesis that the coefficients are zero. The null hypothesis that these coefficients are jointly zero is rejected in almost all cases at the 1% level and in one case at 5% level. Hence, aggregate liquidity as a whole is a price variable in equity markets.

The panel data used consists of data on only 30 stocks from 1997 through 2002, and does not include stock market data from the 2008 financial crisis. In order to see if there are any significant changes to the results from this, we collect daily stock level data spanning over a decade from 1997 through 2008 for over 2000 stocks. Thus, we now have data on the 2008 crisis and the relatively calm periods pre-2008. A detailed description of this dataset can be found in Chapter 2 of this dissertation. Running an ARCH model on a large panel data such as this would be difficult since the maximum likelihood estimation methodology is sensitive to the number of panels and length of the time period. We split this large sample in to four samples according to a stock's average market capitalization over this time period. [12] The model of interest is:

$$R_{it} = \alpha_i + \beta_i^{Mkt}\, MKT_t + \beta_i^{Smb}\, SMB_t + \beta_i^{Hml}\, HML_t + \beta_i^{L}\, L_t + \pi_1\, LiqBeta_{it} \text{ x } L_t + \pi_2\, Spread_{it} \text{ x } L_t + \pi_3 Volume_{it} \text{ x } L_t + \pi_4\, Shrout_{it} \text{ x } L_t + \epsilon_{it}$$

where *LiqBeta* is the beta on a stock obtained from the regression of its return on the market liquidity measure, *Spread* is the effective spread of a stock, *Volume* is the dollar volume of the stock, and *Shrout* is the number of shares outstanding for that stock. Just like before, the stock characteristics used in the regressions are six-month rolling averages of that particular characteristic. In order to obtain quick convergence, the Fama-French factors and the market liquidity variable are regressed out from both sides of the above returns equation. Table 11 shows the results from the following regression for each sample:

$$R_{it+1}^{E} = \alpha_i + \pi_1\, LiqBeta_{it}^{E} \text{ x } L_t + \pi_2\, Spread_{it}^{E} \text{ x } L_t + \pi_3 Volume_{it}^{E} \text{ x } L_t + \pi_4\, Shrout_{it}^{E} \text{ x } L_t + \epsilon_{it}$$

$$\epsilon_{it} = e_{it} * \sigma_{it}, where\ e_{it} \sim N(0,1)$$

$$\sigma_{it}^2 = \delta_0 + \delta_1 \epsilon_{it-1}^2 + \delta_2 \epsilon_{it-1}^2 \mathbf{1}_{\boldsymbol{\epsilon_{i,t-1}>0}}$$

We see that the liquidity beta of a stock is statistically significant for the micro-cap and large-cap samples. The coefficient on liquidity beta in these two samples has a negative sign implying that, on average, high liquidity beta stocks have lower expected returns. The effect found here is opposite to the effect found in Pastor and Stambaugh (2003) or Acharya and Pedersen (2004.) These papers find that on average, stocks with high return sensitivity to market liquidity (or a high liquidity beta) have higher expected returns. This is because investors would not require a premium to hold a stock whose return is high when market

[12] Stocks with average market capitalization lower than $250 million are referred to as 'micro' stocks, those with average market capitalization between $250 million and $2 billion are called 'small' cap stocks. 'Medium' refers to stock with average market capitalization between $2 billion and $5 billion while 'Large' refers to stocks with average market capitalization of $10 billion and above.

liquidity is low. One reason for why we find the opposite effect could be because our dataset includes the market downtown of 2008. It is plausible that given market liquidity was very low during that period, liquidity demanders sought out stocks whose returns were not influenced by market events. We believe this finding emphasizes the importance of controlling for economic conditions that can influence the relationship between the sensitivity of an asset's return to market liquidity and its expected returns. Varying the data window and running the ARCH regressions for each of these windows would be an interesting (though time-consuming) exercise that would shed some light on this hypothesis.

1.5 Conclusion

This paper analyzes how a stock's return sensitivity to aggregate market liquidity (also called comovement with market) and its own individual level of liquidity affects its returns, over and above what volatility (both stock-level and aggregate) can explain. The answer to this question, as posed in the introduction, is that both the level of stock liquidity and its comovement matters when we want to analyze the effects of liquidity on stock returns. We assume that a stock's sensitivity to the market liquidity factor is time-varying and that stock-level characteristics such as volume, effective spread, average historical returns, historical liquidity beta, and number of shares outstanding can be used in modeling the time-varying behavior of the liquidity betas. This assumption is substantively proven by regressing stock returns on stock and market liquidity and volatility variables and testing the validity of the assumption through F-tests. Stock level conditional variance calculated through GJR-ARCH model helps in controlling for stock level risk. Going little further than the past literature, controlling for market level risk helps in identifying crisis periods in this duration because a distressed economic scenario affects riskiness and the liquidity aspect of stocks simultaneously.

Furthermore, past literature adheres to using a single holding period return, say daily or monthly. But liquidity affects investor's preferences depending on their holding period; that is, liquidity might not be as important a state variable for a long term investor as it is for the short term investor. Hence, the second main question of whether these holding period returns are affected differently is also studied. We find that for weekly holding period returns, average dollar volume, after controlling for market liquidity, is consistently significant with an average coefficient of 2.3%. For monthly holding period, both average dollar volume and average number of shares outstanding are consistently significant. The coefficient on average volume in this case lies between 8.7 and 12.6% and that of number of shares outstanding lies between -7.4% and -17%. Average market variance, level of individual stock variance (ARCH-in-mean), and the Fama-French factors are used as control variables and are significant. Average effective spread is also used as a stock specific factor to predict returns, but it is not consistently significant across specifications. However, the factors used for time-varying beta on market liquidity are all jointly significant (i.e. not equal to zero) in most of the regressions. The joint hypothesis test to determine if aggregate market liquidity is a priced variable rejects the null that coefficients are jointly zero at 1% level in all of the cases.

Many further extensions are possible in this line of research. Only weekly and monthly holding periods are constructed and analyzed at the market micro-structure level in this paper. Thus, data leading up to the liquidity crisis can be modeled with the usage of higher holding periods (quarterly and annual) and usage of models with macroeconomic factors. Using macroeconomic factors would also help in obtaining a time-varying market price of risk, as opposed to the constant price of risk at the market microstructure level used in this paper. Econometrically, using state-space model representation for the analysis in this paper might be more intuitive and a better way to model time-varying liquidity.

Table 1.1 : Excess Returns

Variable	Obs	Mean	Std. Dev	Min	Max
Excess Returns From CRSP Daily Returns					
Daily	42,007	0.06	3.03	-39.02	44.98
Weekly	41,891	0.31	6.64	-59.08	65.95
Monthly	41,456	1.2	12.73	-95.32	90.6
Excess Returns From WRDS Intra-Day Sampling					
Daily	42,855	0.01	2.91	-77.85	77.8
Weekly	42,735	0.04	6.65	-82.23	75.31
Monthly	42,225	0.13	13.47	-108.89	82.4
Excess Returns (between -50% and +50%) From WRDS Intra-Day Sampling					
Daily	42,852	0.01	2.84	-44.03	35.24
Weekly	42,710	0.06	6.48	-49.6	48.99
Monthly	42,222	0.14	13.46	-108.89	82.4

Table 1.2: Summary Statistics

Variable	Obs	Mean	Std. Dev	Min	Max
Liquidity Beta	39,285	0.43	1.85	-13.95	22.31
Volume ($'million)	42,856	212.58	577.79	0	15494.39
Shares Outstanding ('000s)	42,856	673.62	1251.51	2.88	6730

Table 1.3: Table of Correlations

	Liquidity Beta	Effective Spread	Gamma	Volume	Return	Shr. Out.
Liquidity Beta	1					
Spread	-0.09	1				
Gamma	-0.11	0.27	1			
Volume	0.08	-0.24	-0.01	1		
Return	-0.08	-0.07	-0.11	-0.01	1	
Shr. Out.	0.07	-0.27	-0.01	0.85	-0.05	1

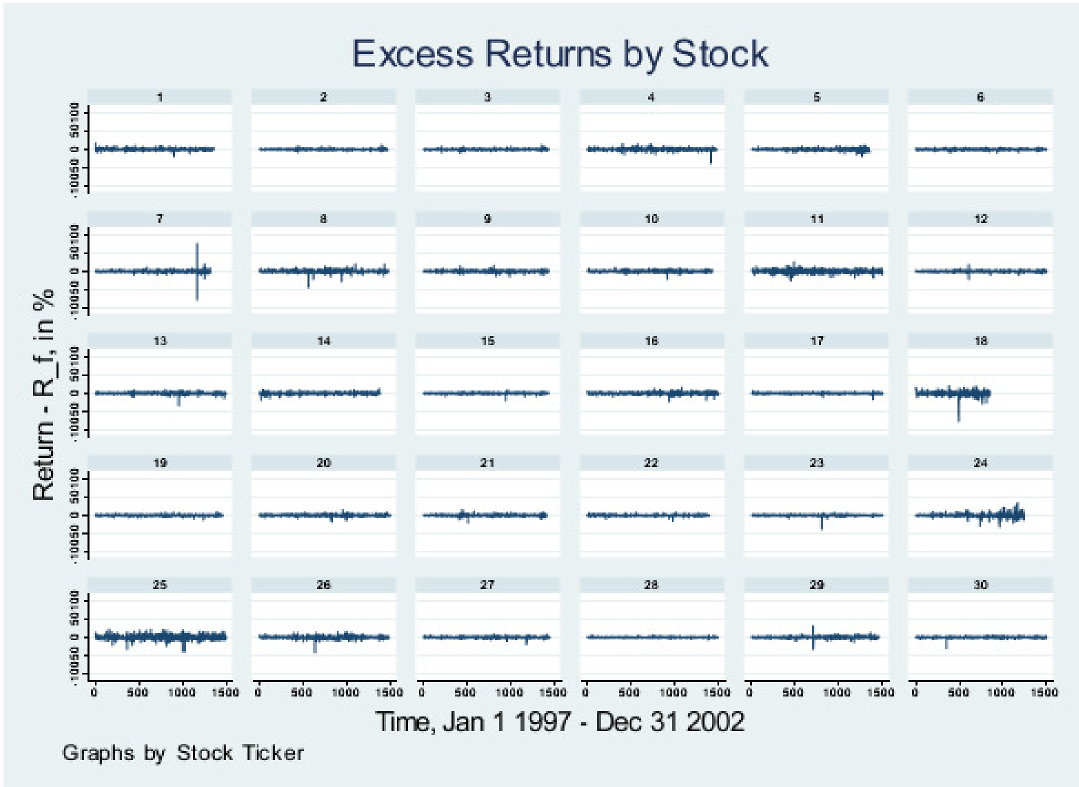


Figure 1.1: Excess Returns by Stock

Table 1.4: Pastor-Stambaugh Gamms by Stock

Ticker	N	Mean	Std. Dev	Max	Min	Skewness
1	1362	1.1420	4.8501	20.4965	-11.8487	1.3331
2	1470	-0.0023	0.0034	0.0041	-0.0143	-1.4394
3	1437	-0.0400	0.1270	0.4519	-0.4029	0.4951
4	1481	-0.0658	0.5417	1.4652	-1.8123	0.0563
5	1350	-2.7812	8.6473	21.8945	-28.6661	-0.4389
6	1495	-0.0024	0.0044	0.0091	-0.0121	0.3745
7	1323	-0.3887	2.9222	18.4118	-8.8695	2.1673
8	1484	0.1273	1.7479	11.0964	-5.1650	3.5406
9	1432	-0.0265	0.0723	0.1585	-0.2265	-0.3759
10	1429	-0.0113	0.0474	0.1757	-0.1614	0.0125
11	1498	-0.0344	0.1792	0.4907	-0.7795	-0.9644
12	1498	-0.0025	0.0035	0.0047	-0.0148	-0.6860
13	1502	-0.0030	0.0037	0.0062	-0.0135	-0.7101
14	1385	-0.1416	0.5945	2.9525	-1.3988	1.7956
15	1437	-0.0164	0.0673	0.0788	-0.4588	-3.7571
16	1504	-0.0005	0.0006	0.0010	-0.0036	-1.3762
17	1501	-0.0014	0.0023	0.0040	-0.0065	-0.3836
18	849	6.4935	100.0672	379.5638	-366.2490	-0.0178
19	1467	-0.0009	0.0018	0.0038	-0.0043	0.4785
20	1505	-0.0004	0.0004	0.0009	-0.0015	-0.0987
21	1414	-0.0072	0.0257	0.0715	-0.0709	0.2736
22	1398	-0.0918	1.8982	7.5217	-7.6412	-0.2655
23	1502	-0.0005	0.0028	0.0070	-0.0082	-0.1364
24	1244	-0.0105	0.8258	2.8476	-3.0950	-0.4112
25	1503	0.0462	0.3479	2.1786	-0.9485	3.1800
26	1487	-0.0637	0.3467	0.8958	-1.2590	-1.2826
27	1450	-0.0049	0.0084	0.0157	-0.0271	-0.0363
28	1498	-0.1462	0.4217	1.2129	-0.9009	0.4878
29	1455	-0.1881	1.4673	7.6287	-4.0692	2.4031
30	1496	-0.0232	0.0663	0.1669	-0.2492	-0.2768

Table 1.5: Pooled Regression Results with Six-month Averaged Stock Characteristics

	Weekly	**Monthly**
$\boldsymbol{L_t}$	1.304	5.209
	[0.649]	[1.995]
$\boldsymbol{LiqBeta_{it-1}}$ **x** $\boldsymbol{L_t}$	0.1106	1.127
	[0.953]	[1.86]
$\boldsymbol{Spread_{it-1}}$**x** $\boldsymbol{L_t}$	1.243	0.356
	[1.401]	[1.539]
$\boldsymbol{Volume_{it-1}}$ **x** $\boldsymbol{L_t}$	2.22**	12.63***
	[1.004]	[1.617]
$\boldsymbol{R_{it-1}}$**x** $\boldsymbol{L_t}$	-26.28**	-27.68
	[10.17]	[56.98]
$\boldsymbol{Shrout_{it}^E}$ **x** $\boldsymbol{L_t}$	-6.89	-17.113***
	[4.01]	[2.836]
$\boldsymbol{MKT_t}$	0.637***	0.350***
	[0.0372]	[0.0391]
$\boldsymbol{SMB_t}$	0.345***	0.0868***
	[0.0522]	[0.0559]
$\boldsymbol{HML_t}$	0.569***	0.224***
	[0.0562]	[0.0653]
$\boldsymbol{\sigma_{Mkt,t-1}^2}$	6.751***	20.94***
	[1.512]	[3.007]
$\boldsymbol{\sigma_{it-1}^2}$	0.0021***	0.0075
	[0.0010]	[0.0076]
$\boldsymbol{\epsilon_{it-1}^2}$	0.785***	0.991***
	[0.0264]	[0.0122]
$\boldsymbol{\epsilon_{it-1}^2 1_{\epsilon_{i,t-1}>0}}$	0.132	-0.021
	[0.0654]	[0.0159]

Note: Standard errors are in parentheses. *** *p<0.01*, ** *p<0.05*, **p<0.10*. Only the main variables of interest are reported in the tables.

Table 1.6: Pooled Regression Results with One-month Averaged Stock Characteristics

	Weekly
$\boldsymbol{L_t}$	1.343**
	[0.660]
$\boldsymbol{LiqBeta_{it-1}}$ x $\boldsymbol{L_t}$	-0.026
	[0.819]
$\boldsymbol{Spread_{it-1}}$x $\boldsymbol{L_t}$	1.287
	[0.861]
$\boldsymbol{Volume_{it-1}}$ x $\boldsymbol{L_t}$	0.444
	[1.094]
$\boldsymbol{R_{it-1}}$x $\boldsymbol{L_t}$	3.808
	[7.489]
$\boldsymbol{Shrout_{it}^{E}}$ x $\boldsymbol{L_t}$	1.240
	[1.607]
$\boldsymbol{MKT_t}$	0.646***
	[0.0375]
$\boldsymbol{SMB_t}$	0.365***
	[0.0534]
$\boldsymbol{HML_t}$	0.592***
	[0.0567]
$\boldsymbol{\sigma_{Mkt,t-1}^2}$	7.491***
	[2.049]
$\boldsymbol{\sigma_{it-1}^2}$	0.0025***
	[0.0009]
$\boldsymbol{\epsilon_{it-1}^2}$	0.773***
	[0.0281]
$\boldsymbol{\epsilon_{it-1}^2 1_{\epsilon_{i,t-1}>0}}$	0.147
	[0.0627]

Note: Standard errors are in parentheses. *** *p<0.01*, ** *p<0.05*, **p<0.10*. Only the main variables of interest are reported in the tables.

Table 1.7: UnPooled Regression Results with Six-month Averaged Stock Characteristics

	Weekly	**Monthly**
$\boldsymbol{LiqBeta_{it-1}}$ x $\boldsymbol{L_t}$	1.099	0.2141
	[1.149]	[2.624]
$\boldsymbol{Spread_{it-1}}$x $\boldsymbol{L_t}$	-0.610	3.5932*
	[1.684]	[2.009]
$\boldsymbol{Volume_{it-1}}$ x $\boldsymbol{L_t}$	2.399**	11.41***
	[1.903]	[4.0545]
$\boldsymbol{R_{it-1}}$x $\boldsymbol{L_t}$	-11.41	1.543
	[13.236]	[15.52]
$\boldsymbol{Shrout_{it}^{E}}$ x $\boldsymbol{L_t}$	1.584	-7.48***
	[1.805]	[3.49]
$\boldsymbol{\sigma^2_{Mkt,t-1}}$	4.474***	31.69***
	[1.578]	[3.81]
$\boldsymbol{\sigma^2_{it-1}}$	0.0014***	-0.0024***
	[0.0009]	[0.0004]
$\boldsymbol{\epsilon^2_{it-1}}$	0.800***	1.060***
	[0.0300]	[0.0235]
$\boldsymbol{\epsilon^2_{it-1}1_{\epsilon_{i,t-1}>0}}$	0.0934	-0.159***
	[0.0582]	[0.0394]

Note: Standard errors are in parentheses. *** *p<0.01*, ** *p<0.05*, **p<0.10*. Only the main variables of interest are reported in the tables.

Table 1.8: UnPooled Regression Results with One-month Averaged Stock Characteristics

	Weekly	**Monthly**
$\boldsymbol{LiqBeta_{it-1}}$ **x** $\boldsymbol{L_t}$	1.458	3.028***
	[1.254]	[1.397]
$\boldsymbol{Spread_{it-1}}$**x** $\boldsymbol{L_t}$	-1.539	-11.83***
	[4.041]	[2.2336]
$\boldsymbol{Volume_{it-1}}$ **x** $\boldsymbol{L_t}$	0.0108	9.145***
	[1.242]	[3.314]
$\boldsymbol{R_{it-1}}$**x** $\boldsymbol{L_t}$	-3.458	-16.90***
	[3.396]	[4.352]
$\boldsymbol{Shrout_{it}^{E}}$ **x** $\boldsymbol{L_t}$	3.411**	-7.431***
	[1.630]	[2.515]
$\boldsymbol{\sigma_{Mkt,t-1}^{2}}$	4.643***	30.25***
	[1.622]	[3.927]
$\boldsymbol{\sigma_{it-1}^{2}}$	0.0015	-0.0025***
	[0.0009]	[0.0004]
$\boldsymbol{\epsilon_{it-1}^{2}}$	0.800***	1.060***
	[0.0300]	[0.0495]
$\boldsymbol{\epsilon_{it-1}^{2}\mathbf{1}_{\epsilon_{i,t-1}>0}}$	0.100*	-0.154***
	[0.0590]	[0.0361]

Note: Standard errors are in parentheses. *** *p<0.01*, ** *p<0.05*, **p<0.10*. Only the main variables of interest are reported in the tables.

Table 1.9: UnPooled Regression Results with Six-month Averaged Stock Characteristics (beta on Lt constant)		
	Weekly	**Monthly**
L_t	1.450**	3.155**
	[0.094]	[1.459]
$LiqBeta_{it-1}$ x L_t	-0.0307	4.900***
	[1.049]	[1.286]
$Spread_{it-1}$x L_t	1.3251	-5.967
	[2.818]	[4.220]
$Volume_{it-1}$ x L_t	2.267**	6.924***
	[0.9779]	[1.776]
R_{it-1}x L_t	-23.80***	-10.58
	[9.116]	[23.7730]
$Shrout_{it}$ x L_t	0.088	12.40***
	[1.332]	[2.2855]
$\sigma^2_{Mkt,t-1}$	4.178***	32.11***
	[2.048]	[3.239]
σ^2_{it-1}	0.0011	-0.0025***
	[0.0009]	[0.0004]
ϵ^2_{it-1}	0.791***	1.056***
	[0.0309]	[0.0219]
$\epsilon^2_{it-1}\mathbf{1}_{\epsilon_{i,t-1}>0}$	0.077	-0.164***
	[0.0635]	[0.0321]

Note: Standard errors are in parentheses. *** *p<0.01*, ** *p<0.05*, **p<0.10*. Only the main variables of interest are reported in the tables.

Table 1.10: UnPooled Regression Results with One-month Averaged Stock Characteristics (beta on Lt constant)

	Weekly	**Monthly**
L_t	1.673**	3.301**
	[0.742]	[1.095]
$LiqBeta_{it-1}$ x L_t	1.011	2.110***
	[0.988]	[1.135]
$Spread_{it-1}$x L_t	0.883	-12.87***
	[3.924]	[1.494]
$Volume_{it-1}$ x L_t	0.2421	6.131***
	[0.814]	[1.0043]
R_{it-1}x L_t	-6.251***	-16.43***
	[2.815]	[4.92]
$Shrout_{it}$ x L_t	0.9911	-12.40***
	[1.378]	[1.929]
$\sigma^2_{Mkt,t-1}$	4.641*	32.63***
	[2.471]	[2.754]
σ^2_{it-1}	0.0004	-0.0025***
	[0.001]	[0.0004]
ϵ^2_{it-1}	0.796***	1.067***
	[0.0299]	[0.0202]
$\epsilon^2_{it-1}\mathbf{1}_{\epsilon_{i,t-1}>0}$	0.055	-0.178***
	[0.0637]	[0.0361]

Note: Standard errors are in parentheses. *** *p<0.01*, ** *p<0.05*, **p<0.10*. Only the main variables of interest are reported in the tables.

Table 1.11: ARCH Regression of Returns on Liquidity Determinants				
	MICRO	**SMALL**	**MEDIUM**	**LARGE**
$LiqBeta_{it}^{E}$ x L_t	-.4892*	-.2363	-0.0206	-1.6223**
	[.2783]	[.3576]	[.3800]	[.6753]
$Volume_{it}^{E}$ x L_t	-9.72E-08	2.53e-08	7.06E-09	5.70e-10
	[4.13e-08]	[2.63e-08]	[6.12e-09]	[7.78e-10]
$Shrout_{it}^{E}$ x L_t	1.77E-06	5.10e-08	2.69E-08	1.70e-08
	[1.11e-06]	[6.74e-07]	[2.34e-07]	[2.15e-08]
$Spread_{it}^{E}$	0.2783	.1112	-0.0491	.2616
	[.1315]	[.2589]	[.1947]	[.3401]
ϵ_{it-1}^{2}	.53443***	.5471***	.6381***	.5658***
	[.0147]	[.0160]	[.0190]	[.0261]
$\epsilon_{it-1}^{2}\mathbf{1}_{\epsilon_{i,t-1}>0}$	0.0118	-.0591***	-.1450***	-.0834***
	[.0235]	[.0193]	[.0220]	[.0311]
Observations	633,202	639,433	603,027	400,613

Note: Each column is a separate regression. Stock fixed effects and year, month, and day fixed effects are included. Robust standard errors are in parentheses.

1.6 References

Amihud, Yakov, and Haim Mendelson, 1986, "Asset pricing and the bid-ask spread," *Journal of Financial Economics*, 17.

Amihud, Yakov, and Haim Mendelson ,1991," Liquidity, maturity, and the yields on U.S. Treasury securities," *Journal of Finance*, 46(4).

Amihud, Yakov, Haim Mendelson, and Lasse H. Pedersen, 200, "Liquidity and Asset prices," *Foundations and Trends in Finance*, 1:4.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam, 1998, "Alternative Factor Specifications, Security Characteristics, and the Cross-Section of Expected Stock Returns," *Journal of Financial Economics*, 49.

Brennan, Michael J., and Avanidhar Subrahmanyam, 1996, "Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns," *Journal of Financial Economics*, 41.

Campbell, John Y., Sanford J. Grossman, and Jiang Wang, 1993, "Trading Volume and Serial Correlations in Stock Returns," *The Quarterly Journal of Economics*, 108(4), pp. 905-939.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, "Commonality in Liquidity," *Journal of Financial Economics*, 56.

Ding, Zhuanxin, Clive W.J. Granger, and Robert F. Engle,1993, "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1(1).

Engle, Robert F.,1982, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50(4).

Fama, Eugene F., and Kenneth R. French. 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1).

Glosten, Lawrence R., Ravi Jagannathan., and David E. Runkle, 1993, "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The Journal of Finance*, 48(5).

Hasbrouck, Joel., and Seppi, D.J."Common Factors in Prices, Order Flows, and Liquidity," Journal of Financial Economics, 59.

Huberman, Gur, and Dominika Halka, 2001, "Systematic Liquidity," *Journal of Financial Research*, 24.

Lo, Andrew W., and Jiang Wang, 2000, "Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory," *Review of Financial Studies*, 13.

Maddala, G.S. and Shaowen Wu, 1999, “A Comparative Study of Unit Root Tests With Panel Data and A New Simple Test,” *Oxford Bulletin of Economics and Statistics* 61, 631-652.

Nelson, David B.,1991, "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59.

Pástor, Lubos, and Robert F. Stambaugh, 2003 "Liquidity risk and expected stock returns," *The Journal of Political Economy*, 111(3).

Shanken, Jay., 1990. "Intertemporal asset pricing : An Empirical Investigation," *Journal of Econometrics,* vol. 45(1-2), pages 99-120

Silber, William L., 1991 " Discounts on restricted stock: the impact of illiquidity on stock prices," *Financial Analysts Journal* 47(4).

2 Liquidity and Asset Returns: Evidence from Stock Market Decimalization

2.1 Introduction

Standard asset pricing models focus on identifying relevant factors that can be used to explain and forecast asset returns, and the liquidity of an asset has been found to be one important determinant. Liquidity is defined as the ease with which an asset can be bought or sold, and for common stocks is often measured by the bid-ask spread. The need to model asset prices as a function of liquidity has become an important consideration in the wake of recent crises where liquidity dried up quickly, such as the financial crisis of 2008, the Internet stock bubble crash of 2000, and the Russian debt crisis of 1998, and. Despite the importance of liquidity in determining asset returns, existing approaches have neglected that liquidity is an endogenous phenomenon that is jointly determined with asset prices. The main contribution of this paper is to account for the endogeneity of liquidity in a standard asset pricing model.

Existing research on the relationship between liquidity and asset returns assumes liquidity to be an exogenous phenomenon, but liquidity can be influenced by many factors which themselves affect stock returns, such as macroeconomic conditions or past stock market performance. As a result, loss of liquidity needs to be incorporated in to models of financial assets so as to forecast returns correctly. To identify the effect of endogenous liquidity in an asset returns equation, we propose an instrument based on the U.S. stock market decimalization. The "decimalization" program was undertaken by the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), and the National Association of Securities Dealers Automated Quotations (NASDAQ) exchanges over the years 2000 and 2001. Under this program, the tick size of each stock—the smallest increment by which price could change--was lowered for all stocks from one-sixteenth of a dollar to one-hundredth of a dollar. Starting in August 2000 and reaching completion in March 2001, the program was phased in over a period of several months across the three exchanges. Each exchange had its own decimalization phases with different sets of stocks adopting the lower tick size in different phases. The move towards decimalizing stock quotes changed a stock's liquidity directly by narrowing the bid-ask spread, but was plausibly unrelated to stock returns. [13] While many researchers have analyzed the effects of decimalization (e.g., Goldstein and Kavajecz, 2000; Bessembinder, 2003), it has not been used as an instrument for stock returns. [14]

Asset liquidity can be thought of as a confluence of three notions: spread, depth, and resiliency. An asset is considered liquid if the difference in bid and offer prices, known as the spread, is not large relative to its price. The market for the asset is deep if there are frequent buy and sell orders, and an asset is resilient if the price of the asset can quickly return to fundamental levels after large buy or sell orders. Current economic conditions can systematically affect liquidity across stocks. For example, in Brunnermeier and Pedersen (2009), during economic downturns market makers or liquidity suppliers can face funding constraints that causes them to reduce their market making activities, thus inducing low levels of liquidity across all stocks in the market. [15] Poor economic conditions can also influence the behavior of stock returns. Hamilton and Lin (1996) find that stock market returns vary over time and that economic downturns can make returns more volatile. Campbell, Gross, and Wang (1993) motivate a model of risk-averse market makers who have to accommodate buying or selling pressures from "liquidity" or "noninformational" traders. By definition, such

---

[13] The decimalization program was put in to place based on the Common Cents Stock Pricing Act of 1997 whose aim was to "…eliminate legal impediments to quotation in decimals for securities transactions in order to protect investors and to promote efficiency, competition, and capital formation." The text of the original bill can be found at http://www.gpo.gov/fdsys/pkg/BILLS-105hr1053ih/pdf/BILLS-105hr1053ih.pdf.

[14] A more general relationship between reduced transaction costs and economic growth has been advocated in Bencivenga, Smith, and Starr (1995, 1996).

[15] Eisfeldt (2004) is another recent theoretical model of liquidity that is endogenously determined as a function of productivity. When productivity is high, investors initiate risky projects as their expected return is high and hold very little riskless assets. Since the outcome of risky projects affects an investor's income, only claims to high-quality projects are sold so as to convert these assets in to current income, thus increasing market liquidity.

traders could be facing wealth constraints that force them to liquidate their assets and we can expect an over-whelming presence of such traders during economic downturns. For reasons posited here, we see that economic conditions can affect both stock liquidity and stock returns. Hence, it is necessary to address the joint determination of stock returns and liquidity when attempting to estimate the effect of liquidity on stock returns.

Empirical modeling of asset liquidity as an endogenous variable is a relatively new endeavor. Some recent studies have used market-level variables to explain individual stock liquidity, but have been unable to identify its direct effect on asset returns. Acharya and Pedersen (2005) develop a standard asset pricing model with liquidity and examine several channels through which liquidity can affect asset prices. Apart from the usual channels of sensitivity of stock liquidity to market liquidity and the sensitivity of stock return to market liquidity, they study a new channel that had not been considered in the previous literature; stock liquidity sensitivity to market return. While they find this channel to be a significant factor in explaining asset returns, they do not study endogeneity in stock liquidity. The main idea of market factors affecting liquidity put forth in this paper is shown empirically by Hameed, Kang, and Viswanathan (2010), henceforth HKV (2010), who find that stock market declines decrease stock liquidity, especially during times of tightness in funding market. However, they are only able to indirectly show the relationship between endogenous liquidity and asset returns through several trading strategies.

This paper explicitly models endogenous liquidity in a standard asset pricing framework. Decimalization status of a stock is used as an instrument in a two-stage least squares (2SLS) framework to identify the effect of endogenous liquidity on asset returns. First stage results confirm the relevance of this instrument. Depending on the sample period, decimalization leads to a 0.1 to 0.7% decrease in effective spreads, the measure of illiquidity used here. By modeling the endogeneity of liquidity, we find that increasing the illiquidity of a stock leads to higher asset returns. In a 2SLS model with stock fixed effects, the effect of illiquidity on asset returns is about 1.2% per year. This result is consistent with the theoretical predictions of Acharya and Pedersen (2005) that as a stock becomes more illiquid, its future expected returns increase. The 2SLS estimates are orders of magnitude higher than the pooled least squares annualized effect of 0.11%. The large difference between OLS and 2SLS estimates demonstrates the importance of incorporating the endogeneity of liquidity. These results are robust to alternate model specifications.

Understanding and accounting for asset liquidity is a crucial step in valuing any asset. The recent financial crisis has shown that failing to account for asset liquidity can lead to major pricing errors. The following quote from Federal Reserve Chairman Ben Bernanke speaking on the financial market turmoil underscores the importance of liquidity risk management:

> Some more-successful firms also consistently embedded market liquidity premiums in their pricing models and valuations. In contrast, less-successful firms did not develop adequate capacity to conduct independent valuations and did not take into account the greater liquidity risks posed by some classes of assets.
>
> -Ben Bernanke, May 15, 2008

The analysis in this paper focuses on the commonality and time-variation in asset liquidity, and is a first step in empirically tying together endogenous liquidity with standard asset pricing models.

The outline of the rest of the paper is as follows: Section II focuses on the basic model and pins down the reasons behind endogenous liquidity; Section III contains a description of the data; Section IV contains discussion of regression results and some robustness checks; and Section V concludes.

2.2 Model

The basic model of illiquidity and stock returns is written down as follows:

$$E_t\left(R_{it+1}^E\right) = \alpha + \beta ESPR_{it} + \boldsymbol{\delta}\mathbf{X_t} + \epsilon_{it+1}\,, \qquad (1)$$

where R_{it}^E is the daily return in excess of the return on market and two additional market factors based on stock size and book-to-market ratio proposed in Fama and French (1993), $ESPR_{it}$ is relative spread, a measure of stock illiquidity, and $\mathbf{X_t}$ is a vector of control variables such as innovations in aggregate market liquidity as developed in Pastor and Stambaugh (2003) and either the changes in the daily business index of Aruoba, Diebold, and Scotti (2009) or lags of stock and market returns. A detailed description of these variables can be found in the Data and Appendix sections. The coefficient of interest is β, which gives the effect of illiquidity on returns one day ahead. As long as an increase in stock illiquidity increases its future expected returns, the expected sign of β is positive. This relationship can be motivated either by the demand for liquidity from investors or the supply of liquidity by market makers. With the help of indirect evidence obtained on financial intermediaries, HKV (2010) empirically test the collateral view of liquidity whereby market makers, as suppliers of liquidity, charge a premium for market-making activities after stock market declines, especially when such market declines affect their funding or collateral constraints. Thus, we would expect to see an increase in returns after huge market downturns. Using short-term price reversals as a measure of the return to supplying liquidity and counter-trend trading strategies, they estimate that liquidity suppliers demand a return of 1.18% per week conditional on large negative market returns. This is much higher than the 0.58% per week that they would unconditionally demand. From the perspective of investors, short term investors value liquidity if they want to be able to liquidate an asset quickly and without incurring much loss. Thus, they would demand a lower price today or expect higher returns in the future to hold an illiquid asset in their portfolios. In their seminal paper, Amihud and Mendelson (1986) model asset prices as a function of the spread and find a positive relationship between spread and returns. Acharya and Pedersen (2005) calibrate the average turnover for stocks and estimate the return premium due to the average level of illiquidity across portfolios to be 3.5% per annum.

Most of the literature assumes stock illiquidity to be exogenous and estimates some form of equation (1). However, recent empirical studies document considerable variation over time and commonalities in liquidity. Commonality refers to the notion that there is (are) an underlying factor(s) that affect liquidity of all stocks at a certain point in time. For example, using latent factor models, Korajczyk and Sadka (2008) document significant commonalities across assets for many measures of liquidity, and find common factors that are correlated across measures of liquidity. They extract a common, underlying liquidity factor and find that shocks to this factor can be explained by past return shocks. Other studies such as Chordia, Roll, and Subrahmanyam (2000), Huberman and Halka (1999), and Hasbrouck and Seppi (2001) demonstrate commonalities in liquidity measures, but do not explicitly study the sources of such commonalities or the resulting effect on returns. Chordia, Roll, and Subrahmanyam (2000) posit that changes in trading volumes affecting optimal dealer inventory levels or asymmetric information within industries can be some reasons for the existence of liquidity commonalities. Vayanos (2004) develops a model of time-varying liquidity premia and theorizes that investors are essentially fund managers whose willingness to hold illiquid assets decreases during volatile times in the market.

Empirical testing of how market factors affect liquidity are sparse, but HKV (2010) find that a decline in market returns and an individual stock's return lead to a decline in that stock's level of liquidity, especially during times of tightness in funding markets. [16] Such market declines affect both the level of liquidity and commonalities in liquidity. They also document an asymmetric effect in that stock market declines affect liquidity much more than stock market increases. Many theoretical models explain the relationship between market declines and liquidity differently. But the basic idea is that large market declines increase the demand for liquidity as investors liquidate their positions across many assets.[17] At the same time, the supply of liquidity declines as market-makers, dealers, and brokers (suppliers of liquidity) hit their wealth or funding constraints. HKV (2010) empirically find that supply effects of liquidity are more significant than demand effects in explaining changes to stock liquidity.

Another source of endogenous liquidity can be the standard business cycle. Eisfeldt (2004) puts forth a theoretical model of liquidity that is endogenous and varies positively with productivity. She defines liquidity as the ability to convert to current income the value of future expected payoffs from long-term assets. During times of high productivity, the relative return on risky investments can be high and many such projects are sold in order to supplement current income for use in consumption and new investment. Thus, market liquidity increases under this scenario and liquidity is pro-cyclical.

In this paper, we estimate the following basic model:

$$E_t\left(R^E_{it+1}\right) = \alpha + \beta\, ESPR_{it} + \boldsymbol{\gamma}\mathbf{X_t} + \epsilon_{it+1}\,, \qquad (2a)$$

[16] Pastor and Stambaugh show that their measure of market liquidity is also affected by stock market declines, but do not account for this relationship in their asset pricing regressions.
[17] Kyle and Xiong (2001), Gromb and Vayanos (2002),Anshuman and Viswanathan (2005), Brunnermeier and Pedersen (2009), Garleanu and Pedersen (2007), and Mitchell et al. (2007)

$$ESPR_{it} = \delta + \boldsymbol{\kappa}\mathbf{X_t} + \mu Z_t + \nu_{it}, \qquad (2b)$$

where R_{it}^{E} is the daily return in excess of the return on market and two additional market factors based on stock size and book-to-market ratio proposed in Fama and French (1993), $ESPR_{it}$ is relative spread, a measure of stock illiquidity, and $\mathbf{X_t}$ is a vector of control variables such as innovations in aggregate market liquidity as developed in Pastor and Stambaugh (2003) and either the changes in the daily business index of Aruoba, Diebold, and Scotti (2009) or lags of stock and market returns. Variables such as past market returns and productivity in the economy can affect both liquidity and future returns. Thus, to solve the problem of identification, an instrumental variable Z_t based on the NYSE and NASDAQ decimalization program is used. Decimalization refers to the change of tick sizes, the minimum price allowable price increment, from $1/16^{th}$ of a dollar to $1/100^{th}$. Stocks were decimalized at different times based on their assigned decimalization phase, as summarized in Table 1. Studies by Goldstein and Kavajecz (2000), Bessembinder (2003) show that decimalization program led to a decrease in the difference between the bid and offer quotes and decreased the depth or the number of shares offered at each quote. Depth or the number of shares at each bid and offer quote decreased as investors did not have to trade large quantities to overcome high transactions costs anymore. As is evident from documents released by the U.S. Congress and the Securities and Exchange Commission, this program was carried out to update tick sizes in the U.S. to that of world financial markets and to enhance liquidity.[18] There is no evidence to suggest that the instrument is correlated with other variables that can explain returns.

2.3 Data

[18] See related documents available online at http://www.sec.gov/hot/decimal.htm.

Data for this paper are obtained from the Wharton Research Data Services (WRDS) database. The sample period is from the beginning of 1997 through the end of 2008. A sample of 2084 U.S. common stocks (share code 10 or 11) listed on the NYSE, NASDAQ, and AMEX are included.[19] The following subsections document the construction of daily stock returns, the average effective spread, and the instrument with information obtained from various data sources. See Appendix for a discussion of other variables used, such as the Fama and French (1993) factors, the market liquidity measure in Pastor and Stambaugh (2003), and the daily business indicator from Aruoba, Diebold, and Scotti (2009).

2.3.1 Daily Prices and Returns

Intra-day data on prices are obtained using WRDS's Trade and Quotes (TAQ) database. Prices are sampled at ten-minute intervals between 10:00am and 3:00pm. The half-hour after market open and the hour before market close are excluded because prices tend to behave erratically during these times.[20] For a stock on a given day, it is possible to have 31 data points. Daily price P_{it} is obtained as the average of the intra-day price data. Keeping in line with the literature, only stocks with price lower than $1000 are included. The daily return for stock *i* on day *t* is calculated as follows:

$$R_{it} = ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right)$$

2.3.2. Effective Spreads

[19] This excludes American depositary receipts, units, shares of beneficial interest, companies incorporated outside the United States, Americus Trust components, close-ended funds, preferred stocks, and real estate investment trusts.

[20] Chan, Christie, and Schultz (1995) document intraday patterns in spreads for NASDAQ securities while Wood and Ord (1995) and Brock and Kleidon (1992) document U-shaped intraday spreads for NYSE securities.

Intra-day bid and offer quotes are obtained in the same way as that of the stock price. The effective spread is defined as twice the absolute difference between the actual trade price and the midpoint of the bid-offer spread. Since quote updates from a trade are frequently recorded ahead of the trade, we cannot match the current trade price with the current quote. The trade price is matched with the quote in effect 5 seconds before the trade time-stamp as advocated in Lee and Ready (1991). Using these quotes, the relative effective spread for a given stock *i* at time *s* on day *t* is:

$$Relative\ Effective\ Spread_{its} = \frac{2 \times \left| P_{its} - \left(\frac{Ask_{its} + Bid_{its}}{2}\right) \right|}{\left(\frac{Ask_{its} + Bid_{its}}{2}\right)}$$

Effective spreads are a better proxy for liquidity than quoted spreads (the difference between bid and offer prices) for several reasons. Lee and Ready (1991) show that 30% of trades occur inside the spread and quoted spreads cannot capture the effect of such trades. Moreover, quotes are often "indications" to trade and may not be firm offers of a trade. Thus, effective spreads better captures the cost of a round-trip order by accounting for any price movements caused by market makers' arrival to fill the order at a better price. To calculate the spread as a percentage of the underlying stock price, we divide by the bid-offer midpoint or the price.[21] Anomalous observations that satisfied the following conditions are deleted[22]:

1. Quoted Spread = Ask – Bid < 0;
2. Effective Spread / Quoted Spread > 4;

[21] Dividing either by price or the bid-offer midpoint did not affect the results. Expressing spread as a percentage can be useful to compare spreads across stock with different trading prices.
[22] Such filters are used in the empirical literature (see, for e.g.,Chordia, Sarkar, and Subrahmanyam (2005)) to remove the effects of incorrect data entries.

Relative effective spreads greater than 0.4 are set to equal 0.4 so that results are not influenced by anomalous spreads. Daily effective spread is obtained by averaging over times *s* within day *t*. The empirical analysis uses proportional effective spread which is calculated as the ratio of daily effective spread and midpoint of the bid-ask spread:

$$ESPR_{it} = \frac{1}{n_{it}} \sum_{s=1}^{n_{it}} Relative\ Effective\ Spread_{its},$$

where n_{it} is the number of observations available on day *t* for stock *i*.

2.3.1 Instrument

The NYSE, AMEX and the NASDAQ followed separate schedules of decimalization, as outlines in Table 1A. Between August 2000 and January 2001, the NYSE undertook four phases of decimalization: 7 stocks were converted on August 28, 2000, 57 stocks on September 25, 2000, 94 stocks on December 4, 2000, and rest of the stocks on January 29, 2001. The NASDAQ decimalized between March and April 2001; 15 stocks were converted on March 12, 2001, 199 stocks on March 26, 2001, and the rest on April 9, 2001. The list of stocks in these different phases is obtained from Chakravarty, Harris, and Wood (2002) and from a search for newspaper articles on *Lexis Nexis Academic*. The instrumental variable *decimalization status* for a stock *i* on day *t* takes on the value of 1 if its prices are quoted in decimals, and 0 if not. Table 1A shows the timeline of the decimalization phases and the number of stocks in each phase by exchange. Decimalization status can be effectively used as an instrument for liquidity only if the process of decimalization is exogenous to asset returns, conditional on observables or control variables. That is, observable characteristics of stocks that were decimalized early should not be very different from stocks that were decimalized late. More importantly, mean returns or excess returns (described below) during the pre-decimalization phase for stocks in the early and late decimalized group should be statistically similar. To investigate this issue, Tables 1B and 1C show mean values of return, excess return, effective spread, volume of trades, and market capitalization for early and late decimalized stocks in the pre-decimalization period for NYSE and NASDAQ stocks. Late decimalizers are those NYSE and NASDAQ stocks decimalized on January 29, 2001 and March 26, 2001, respectively and early decimalizers belong to the group of stocks decimalized before these dates. For NYSE stocks, the pre-decimalization period considered is between August 1999 and August 2000. Varying the time period of this phase did not change the results (not reported here). Table 1B shows the results for variables in level and Table 1C reports the mean values for panel-level demeaned variable, i.e. after removing any firm-specific mean effect. There is no significant difference in returns or excess returns across stocks in the early and late phases. For example, in Table 1C, neither the difference in the mean of returns (t=-0.0002) nor difference in mean of excess returns (t=0.69) for early and late decimalized NYSE stocks is statistically significant. In Table 1B, the t-statistic for difference in returns and excess returns is 0.64 and 0.71, respectively. Although mean returns across the early and late decimalized stocks are statistically indistinguishable, several other stock characteristics are different. In Table 1B, the mean effective spread is statistically significant for the two groups (t=-6.01) with the early group having a lower spread. Similarly, large-cap stocks (with mean of $8 billion) are more likely to have been decimalized early than mid-cap stocks (with mean size of $5.6 billion). There is statistical and economic significance in the mean volume measured as number of shares traded across these two groups (t=45.01). The reason these characteristics differ is that U.S. stock exchanges were worried about the transition to decimals affecting market activity, and chose to implement the process in different phases so as to observe the effects of decimalization for a smaller group before implementing it to the market as a whole.[23] Highly traded stocks or high volume stocks were more likely to have been picked

[23] See citations to Floyd Norris' articles in *The New York Times* in the reference section.

than stocks that traded infrequently so that the effects of decimalization could be easily observed. Moreover, high volume stocks tend to be large firms that have lower spreads, as is observed for early decimalizers. The results for NASDAQ stocks are similar to that of NYSE stocks. The pre-decimalization period considered for NASDAQ stocks is between March 2000 and March 2001. The difference between excess returns, though insignificant at the 10% level, is cause for slight concern (t=-1.56). But when the pre-decimalization period is changed to begin at March 1999, the difference is statistically insignificant (t=0.34). The dot-com bubble burst in March 2000 and this could have influenced some of the results in the middle panel of Tables 1B. Furthermore, to investigate if the significant differences in stock characteristics (other than returns) across the two groups can be captured by a stock specific effect, Table 1C reports the mean across the two groups for all the demeaned stock characteristics; this is obtained by subtracting out the panel-level mean for each stock for each characteristic. Accounting for stock fixed-effects wiped out any significant difference across early and late decimalized stocks for all characteristics. Overall, we see that stock returns were not a significant factor in determining which stocks are decimalized early. Large cap stocks with higher trading volumes and lower effective spreads were most likely to be decimalized early. While the main results pool the stocks across both exchanges, results for NYSE and NASDAQ stocks are separately reported in the section on robustness checks.

2.3.4 Excess Returns

Daily excess returns are calculated for each stock by calculating the residuals from a regression of stock returns (net of the risk-free rate) on the aggregate market return and the Fama French factors obtained from WRDS's Fama French and Liquidity Factors database. A stock's excess return, R_{it}^{E} is obtained as the residual from a stock-by-stock regression of daily return premium in excess of the 1-month Treasury bill rate $(R_{it} - R_{ft})$, on $MKTRF$, the daily value-weighted excess return on all NYSE, AMEX, and NASDAQ stocks minus the one-month Treasury bill rate, *SMB*, the average return on the three smallest (by market capitalization) portfolios minus the average return on the three biggest portfolios, and *HML*, the average return on the value portfolios minus the average return on the growth portfolios. Thus,

$$(R_{it} - R_{ft}) = (\alpha_i + \beta_i^{MKTRF} \times MKTRF_t + \beta_i^{SMB} \times SMB_t + \beta_i^{HML} \times HML_t) + R_{it}^E$$

2.3.5 Summary Statistics

Table 2 gives the summary statistics of the main variables used in this paper. Data is available for 2084 stocks with an average of 1459 days' worth of data for each stock. The average price, unadjusted for splits and dividends, is about \$27, and the average daily return calculated using adjusted prices is 0.012%. Calculating the excess returns by factoring out the Fama-French factors yields an average excess return with mean 0 and standard deviation of 3.65%. In comparison, the average return on the market during this same period is 0.013%. The relative effective spread averaged across stocks is 1.31% and from the decimalization status, we see that the average stock has a decimalized tick size 61% of the time.

2.4 Regression Results

Estimates come from the following OLS panel regression model:

$$E_t\left(R_{it+1}^E\right) = \alpha + \beta ESPR_{it} + \gamma\, \Delta ADS_t + \delta L_t^{MKT} + \epsilon_{it+1}\,, \qquad (1)$$

or the 2SLS panel regression model:

$$E_t\left(R_{it+1}^E\,\right) = \alpha + \beta ESPR_{it} + \gamma\, \Delta ADS_t + \delta L_t^{MKT} + \epsilon_{it+1}\,, \qquad (2a)$$

$$ESPR_{it} = \delta + \mu Z_{it} + \kappa \Delta ADS_t + \phi L_t^{MKT} + \nu_{it}, \qquad (2b)$$

2.4.1 Pooled Regression Results

2.4.1.1 Results for levels of variables

We begin the analysis with the standard OLS regression of excess returns, R_{it+1}^{E}, on the level of illiquidity, measured as the relative effective spread, $ESPR_{it}$. All standard errors are calculated using White's (1980) heteroskedasticity-consistent standard errors. We control for market illiquidity by including Pastor and Stambaugh's (2003) market illiquidity series and the Aruoba-Diebold-Scotti (ADS) business index, a daily index that controls for the general business conditions in the economy. [24] Due to the presence of a unit root in the original series, the ADS business indexed is first-differenced and it is this resulting series that will be used in all further analysis. Table 3 shows the OLS regression of Equation 1. The coefficients on all the right-hand side variables follow the signs usually seen in the literature, and are highly statistically significant. The coefficient on the illiquidity variable is positive because any increase in illiquidity increases expected returns as investors would need to be compensated for holding an illiquid stock. A 1% point increase in stock illiquidity is associated with a 0.013% (1.3 basis points, or *bps*) increase in return at the daily level. Alternatively, a one standard deviation increase in illiquidity (σ=0.029) is associated with a 0.09% per year increase in returns. Positive changes in the ADS business index mean that the economic conditions are improving, and decreases mean that conditions are worsening. Such a betterment of economic conditions could be associated with increase in future market returns and we may expect increased excess returns for stocks that are highly correlated with market returns in sectors such as manufacturing or automobile. Changes in market liquidity are positively correlated with stock returns, consistent with the existing literature on liquidity (Pastor and Stambaugh, 2002 and Chordia, Roll, and Subrahmanyam (2000)). The R^2 of all

---

[24] Appendix section gives more information on these variables and how they are constructed.

these regressions are very close to zero as predicted by the efficient market hypothesis, and the inclusion of time dummies does little to change the magnitude or sign of the coefficients.

Moving on, we instrument for stock illiquidity with the decimalization variable. Results from the two-stage least squares (2SLS) estimation is reported in Tables 4A and 4B. The first stage results in Table 4A show that decimalization status, $Dec\ Status$, is a strong predictor of liquidity: on average, decimalized stocks have a 0.3% to 0.96% lower relative bid ask spread than non-decimalized stocks ($p<0.0001$). The negative sign indicates that average relative spreads decreased in response to decimalization, as expected. To put this value in perspective; the average value of the relative spread in the year before decimalization was 0.014, implying that the average relative spread declined by about 20% (=-0.0028/0.014) in response to decimalization conditional on it being the sole cause of any changes in spread. Previous studies have found comparable increases in liquidity in response to lowered tick sizes (Goldstein and Kavajecz, 2000).

The relationship between changes in the ADS business index and stock illiquidity implied by the results is opposite of theoretical expectations. We might expect a negative sign if positive changes in the index, indicative of bettering economic conditions are associated with a decrease in spreads. But the coefficient in results is positive. Perhaps any change in the business indicator is perceived to be a result of uncertain economic conditions, thus raising spreads or changes in the business conditions only affect stock liquidity through other financial market variables. The coefficient on market liquidity is negative since an increase in market liquidity is associated with a decrease in spreads, especially if the stock's liquidity is correlated with market liquidity. The coefficient signs and magnitudes are robust to inclusion of time dummies, as is seen by comparing them across columns.

The first-stage R^2 is low because the variables included here do not fully explain stock illiquidity. The market-microstructure literature on explaining how and why spreads exist is vast and it is very difficult to obtain data on determinants of liquidity, like measures of adverse selection, market maker inventories, and so on. Additionally, stock illiquidity is persistent, having a first-order autoregressive coefficient of 0.50. Lags of the endogenous regressor are not included as there is only one instrumental variable, and so including them would make the model unidentified. A test of the instrumental variable, decimalization status is reported in Table 4A. Across all regression specifications the instrument is a significant predictor of stock liquidity, satisfying the relevance condition with a F-statistic of at least 3636.73 ($p<0.0001$).

The second-stage results are reported in Table 4B. After accounting for endogeneity in stock illiquidity, the coefficient on stock illiquidity in the second stage has increased to between 0.03% and 0.06%. The coefficients on the business index and market liquidity do not change from the previous OLS regressions and are still significant. Thus, a 1% point increase in spread today increases one-day-ahead returns by about 4 to 6 basis points depending on the specification. On an annual level, this amounts to a 0.28% to 0.40% increase in stock returns.

From the 2SLS regressions results, we see that modeling endogeneity of liquidity is important, especially when we want to study the implications of liquidity on asset prices. The IV estimate (0.40% per year) of the beta of stock illiquidity in the second stage returns equation is orders of magnitude higher than the OLS estimate (0.09% per year), implying that stock illiquidity is negatively correlated to the error term of the returns equation. To see this, consider the following simplified model of the second stage. If y is the daily excess return, x is stock illiquidity, and β_{IV} the coefficient of illiquidity from the 2SLS regression, then β_{OLS} estimates:

$$\hat{\beta}_{OLS} = \frac{dy}{dx} = \hat{\beta}_{IV} + \frac{du}{dx}$$

$$\widehat{\beta}_{OLS} \ll \hat{\beta}_{IV} \rightarrow \frac{du}{dx} < 0$$

The 2SLS results confirm Proposition 2 of Acharya and Pedersen (2005) that conditional expected return increases with illiquidity when illiquidity is persistent. Persistence of illiquidity, as seen here empirically, further implies that contemporaneous illiquidity is negatively correlated with current returns. This is because, when illiquidity is high, the required future return is high. Naturally, prices today must decline, leading to low current return.[25] This proposition is proved in Acharya and Pedersen (2005).

2.4.1.2 Results for demeaned variables

Since we have panel data, to control for unobserved heterogeneity across stocks, all the right-hand side variables are demeaned and the results from the 2SLS regressions with these demeaned variables are reported in Tables 5A and 5B. From here on in text, we will only discuss estimates from the specification with all the time effects included as it is the most complete model. A noticeable change from previous first-stage results is that the magnitude of coefficient on decimalization status has decreased substantially (coefficient = 0.1%), from when variables are not demeaned. However, decimalization status remains highly significant with an F statistic of 456.69 (*$p<0.0001$*). In the second stage, the magnitude of the coefficient on stock illiquidity increases almost ten-fold to 0.20% (20 bps). For a one standard deviation increase in illiquidity (σ=0.024), this is equivalent to an increase in returns of 1.2% (=0.299x0.024x250) per year. Thus, the effect of average correlation within stocks is much stronger than the average correlation across stocks, suggesting that stock illiquidity is

[25] Regressions with price as the dependent variable will be considered for future work.

negatively correlated with the unobserved stock-specific effect. The endogeneity test strongly rejects the null hypothesis that illiquidity is exogenous (*$p<0.01$)*.

Acharya and Pedersen (2005) find that firms with high illiquidity tend to be small (low market capitalization), have high volatility of stock returns, and low turnover of shares. Such firms also trade very infrequently and tend to have fewer days' worth of data. To check if very small firms affected the results, regressions (not reported) were repeated by dropping firms with fewer than 1000 days of daily returns data. Though the coefficient associated with relative spread in the returns equation dropped to 0.11% (11 bps) on average, strength of the instrument variable and the qualitative conclusions about the endogeneity of liquidity both hold.

2.4.1.3 Estimates over various sample periods

The estimates obtained so far are based on a policy change that was carried out over a period of seven months between August 2000 and March 2001. The average effect of the instrument could have been lowered due to the sample period spanning twelve years of daily data. To test the effect of data sample length, the 2SLS regressions were repeated on data samples spanning one through three years around the decimalization policy change. Tables 6A-7 shows regression estimates obtained using 2 and 4 years' worth of daily data. The two year window spans January 2000 to December 2001. There are no substantive changes in the first stage results but the effect of decimalization is higher (0.1-0.7%) than before across most specifications. As expected, the coefficient on the relative spread in the returns equation is higher regardless of whether the variables were demeaned or not. Across specifications in Tables 6B and 7, the effect of illiquidity in returns ranges between 4 and 44 bps. So, a one standard deviation $\sigma = 0.021$ increase in illiquidity raises returns by 0.21% to 2.3% per year.

Results for data spanning six and eight years are not reported. These windows cover the data between January 1998 and December 2003 for the six-year period, and between January 1997 and December 2004 for the eight-year period. These results do not align with that of the other estimates in the paper as the sign of the coefficient on liquidity in the returns equation is negative and statistically significant. While second stage results based on demeaned variables have the correct signs, the first stage results show a positive effect of decimalization status on a stock's illiquidity. Results based on the level of variables have the correct first stage results, but the coefficient on illiquidity is negative in the returns equation. These results could be influenced by the vast variability in stock illiquidity during this period. Indeed, a plot of the weekly illiquidity averaged across stocks show two significant peaks and subsequent drop in the relative spread measure. Including twelve years' worth of data from January 1997 through December 2008 averaged out the effects of these large swings in average stock illiquidity.

2.4.2 Weekly Returns, past return performance, and volatility

Results so far confirm the proposition that liquidity is endogenous and that decimalization is a strong predictor of stock liquidity. However, changes in the ADS business index did not predict changes in illiquidity in the correct direction. Even though the business index is calculated at the daily level, the information used to calculate the index can be a week to two weeks old. We may expect that the stock markets react immediately to any such news about economic conditions, and changes to the ADS business index could only affect stock market illiquidity through changes in volatility of returns. Hamilton and Lin (1996) empirically show this result when they model stock return volatilities over time and find returns to be highly volatile during economic recessions. HKV (2010) also motivate the idea of past stock market declines affecting liquidity. They empirically analyze a weekly model of liquidity to find that past declines in returns lead to higher stock illiquidity and find that this illiquidity effect lasts for two weeks. Hence, we test whether including two week lags of stock and market returns can help predict stock illiquidity in the current week. Owing to the fact that illiquidity is highly correlated with volatility of stock returns, (HKV (2010) and Acharya and Pedersen (2005)), we construct weekly stock volatility σ_{iw} and market volatility σ_{w}^{MKT} as a function of returns, as outlined in French, Schwert, and Stambaugh (1987). All daily variables are also averaged at the weekly level and we have 624 data points for a stock with complete data over the sample period. Also, since the direction of effect between liquidity and volatility is not clear, both current and lagged values of stock and market volatility are included.

The model to be estimated is:

$$E_t\left(R_{i,w+1}^{E}\right) = \alpha + \beta\, ESPR_{iw} + \sum_{k=1}^{2} \gamma_{1k}\, R_{i,w-k}^{E} + \sum_{k=1}^{2} \gamma_{2k}\, R_{i,w-k}^{MKT} + \sum_{k=0}^{1} \gamma_{3k}\, \sigma_{i,w-k} + \sum_{k=0}^{1} \gamma_{4k}\, \sigma_{w-k}^{MKT} + \gamma_5 L_{w}^{MKT} + \epsilon_{i,w+1}\,, \qquad (3a)$$

$$ESPR_{iw} = \delta + \sum_{k=1}^{2} \pi_{1k}\, R_{i,w-k}^{E} + \sum_{k=1}^{2} \pi_{2k}\, R_{i,w-k}^{MKT} + \sum_{k=0}^{1} \pi_{3k}\, \sigma_{i,w-k} + \sum_{k=0}^{1} \pi_{4k}\, \sigma_{w-k}^{MKT} + \pi_5 L_{w}^{MKT} + \mu Z_{iw} + \nu_{iw}, \qquad (3b)$$

Tables 8A through 9B show the results from 2SLS estimation above. Regression specifications are reported with different month and year dummies. Results from the demeaned variables are shown in Tables 9A and 9B. We see that adding the volatility variables are highly informative in explaining stock illiquidity across both columns in Table 9A. All volatility variables have a positive and significant coefficient at the 1% level and the business index is not significant in explaining changes in stock illiquidity. For example in column B of Table 9A, a 1% point increase in current stock volatility is associated with a 0.03% (or 3 bps) increase in stock illiquidity. Spreads are positively affected by return volatility due to higher adverse selection in the market resulting from information asymmetries (Stoll, 1978). At lag one, the effect decreases to 1 bps. Current and lagged market volatility has about a 5 bps effect on stock illiquidity.

Overall, increases in both individual security-level and market-level volatility increases stock illiquidity by 1-4 bps, over and above changes in the business index. These results confirm our hypothesis that changes in the ADS business index can affect stock illiquidity through changes in stock and market return volatilities. The coefficient on lagged stock returns is negative and significant. A 1% point decrease in lagged stock returns increases stock illiquidity by 2.6 to 3.7 bps per week. This is comparable to estimates in HKV (2010). Though the coefficients on lagged market returns are significant, they have the opposite sign. This could be due to high correlation between individual stock returns and the market return at the weekly level. However, results on lagged returns qualitatively confirm the results in HKV (2010) that stock market declines affect stock illiquidity. Decimalization status is still a good predictor of stock illiquidity. A stock's decimalization status has a negative and significant effect on stock illiquidity at least when both month and year time fixed effects are included. When a stock is decimalized, its liquidity increases by approximately 1 bps. Test of the excluded instrument, decimalization status, has a partial F-statistic of 68.46 (*p<0.0001)* and rejects the null hypothesis that the effect of the instrument is zero.

Table 9B shows the second-stage estimates with the variables in levels. We see that the main coefficient of interest, the beta on stock illiquidity, in the returns equation is positive, but only significant when both month and year time effects are included. A one standard deviation increase in illiquidity ($\sigma = 0.0288$) increases returns by 0.72%, (=0.50x0.0288x50) per year. As found in the literature, market liquidity is positively correlated with stock returns. The interpretation of other control variables in the second stage is quite difficult due to collinearity. In columns A and B, the business indicator has a significant negative coefficient implying that expected returns fall when there are positive changes in ADS business index indicative of better economic conditions. One might expect the ADS business index to be most informative when it is used to construct indicator variables for an economy in a recession (ADS<0) and reference points that can account for the severity of that particular recession. [26] An increase in the stock's idiosyncratic volatility is associated with increasing returns, keeping in line with the risk-return relationship, while the effect of market volatility is not robust across specifications. Many of the lagged return and volatility variables are probably correlated with each other and this might hinder separate identification of the coefficients.

To summarize, the results using weekly returns confirm most of the results using daily returns. In the first stage, including past stock and market performance in place of the business indicator is very informative as past individual security-level and market-level returns help predict stock illiquidity. From Tables 8A and 9A, we can see that the relative spread is influenced negatively by past stock returns. The coefficients on market returns are not robust to various time effects, but when they are significant, the negative effect found by other researchers is not seen here as they are probably correlated with stock volatilities. Decimalization status is still significant with a negative effect on illiquidity. The effect of stock illiquidity on stock returns in the second stage is positive and significant in most specifications. There is also evidence of return reversal, as implied by the negative coefficients

[26] Cenesizoglu (2011) empirically shows that stock returns react differently to news about macroeconomic variable depending on whether we are in an expansion or a recession. Also, different firms react to these variables differently in an expansion. We might expect counter-intuitive results of better economic conditions on returns if firms are sensitive to changes in their expected cash flows arising from news about the economy. See Hamilton (2009, 2011) for a detailed discussion on this index and its interpretation.

on past returns. This confirms the collateral view of liquidity that market makers require a higher return to supply liquidity in an uncertain market.

2.4.3 Robustness Checks

2.4.3.1 Price as dependent variable

So far, we have established a positive causal effect of increased illiquidity on one-day ahead stock returns. Since one-day ahead stock returns are calculated as $ln\left(\frac{P_{i,t+1}}{P_{i,t}}\right)$, a related question is whether there exists a negative relationship between illiquidity ($ESPR_{it}$) and stock prices today P_{it}. That is, if a stock becomes illiquid today, would its price fall so as to compensate investors for taking on the liquidity risk? This question is answered by running the 2SLS regression with today's stock price as the dependent variable. Results (not reported here) confirm this hypothesis.

2.4.3.2 Standard errors

Petersen (2009) examines different methods used to calculate standard errors for finance panel datasets in which residuals may be correlated across time or across firms. In the presence of such correlation, it is important to recalculate standard errors allowing for different types of correlation in the error term. Table 10 shows the second-stage results of the 2SLS framework with Newey-West standard errors, which control for heteroskedasticity and autocorrelation up to 11 lags in the daily data sample.[27] The estimates are all significant at the 1% level and the standard errors are

---

[27] The data dependent scheme to select lag length is K = $4\left(\frac{T}{100}\right)^{2/9}$, where T is the number of sample observations over time (Baltagi, 2000). With T = 3020, the nearest K is 9. To be conservative, results are reported in this paper with lag length of 11.

not different from the regressions with White standard errors. The decimalization instrument is significant (F statistic of 136.42) in the first-stage (results not reported).

Table 11 shows the second stage results with clustered standard errors where the clustering is at the firm level. Also called Roger standard errors in the finance literature (Petersen, 2009), clustering allows for arbitrary correlation within each particular cluster. There is evidence of within-firm correlation as the standard errors have increased across specifications. The most complete specification is the last column of Table 11 with all the time dummies included, and it is significant at the 1% level. However, the coefficient on stock illiquidity is now significant at the 10% level for all other specifications. The significance of decimalization instrument is evident from the first-stage (results not reported) F statistic of 12.38. Clustering at the firm level is a conservative approach because it allows the errors in any two firm periods (even years apart) to be correlated.

2.4.3.3 NYSE vs NASDAQ

The empirical finance literature usually considers NYSE and NASDAQ stocks separately since these two exchanges tend to have different trading protocols and many of the NASDAQ stocks are technological firms. A good robustness check would then be to split the sample based on the exchange. Table 12 shows second stage results for NYSE stocks and Table 13 shows result for NASDAQ stocks. Month, year, and day-of-week dummies are included in all specifications. Columns in the table differ by the type of standard errors reported. Substantively, splitting the sample did not change the results and the decimalization instrument is significant (*p-value <0.001).* The coefficient on illiquidity for NYSE stocks is 0.376, so for a one standard-deviation increase in illiquidity (σ=0.0133), returns increase by

0.005% per day or 1.25%, (= 0.376x0.0133x250 trading days) per year. For NASDAQ stocks, though the coefficient is smaller at 0.2065, the standard deviation of the spread variable is higher at 0.036. Thus, the effect on returns of a one standard deviation increase in illiquidity is about 0.0074% per day or 1.85% per year.

2.4.3.4 Fama-French factors in the instrumental variable regression

Excess returns, i.e. the residuals from the regression of daily stock returns on the Fama-French factors, were used as dependent variable in the results discussed so far. However, it is not difficult to imagine a case where the Fama-French factors are correlated with control variables such as the Pastor-Stambaugh market liquidity measure and business conditions index. The 2SLS regressions in (2a) and (2b) are modified as follows, where index *i* refers to the actual stock while index *j* refers to the 3-digit industry code:

$$E_t(R_{it+1}) = \alpha_i + \omega_j^{MKTRF} MKTRF_t + \omega_j^{SMB} SMB_t + \omega_j^{HML} HML_t + \beta ESPR_{it} + \gamma\, \Delta ADS_t + \delta L_t^{MKT} + \epsilon_{it+1}\,, \qquad (4a)$$

$$ESPR_{it} = \delta_i + \psi_j^{MKTRF} MKTRF_t + \psi_j^{SMB} SMB_t + \psi_j^{HML} HML_t + \mu Z_{it} + \kappa \Delta ADS_t + \phi L_t^{MKT} + \nu_{it}, \qquad (4b)$$

MKTRF, SMB, and *HML* are the Fama-French factors, $R_{it+1}$is the daily return of stock *i* on day *t+1*, the (relative) effective spread is given by $ESPR_{it}$, and Z_{it} is the decimalization status of the stock. The fixed effects 2SLS model requires the estimation of many parameters, especially since we cannot constrain the coefficients on each of the Fama-French factors to be the same across stocks. For computational ease and model tractability, the sample is split by market capitalization and the coefficients on the Fama-French factors are estimated for each 3-

digit industry. Tables 14A and 14B shows the main estimates of equations 4a and 4b under the case of robust standard errors. Tables 15A and 15B estimate the same model except standard errors are now clustered by stock. In all the tables discussed below, stocks with average market capitalization lower than \$250 million are referred to as ‘micro’ stocks, those with average market capitalization between \$250 million and \$2 billion are called ‘small’ cap stocks. ‘Medium’ refers to stock with average market capitalization between \$2 billion and \$5 billion while ‘Large’ refers to stocks with average market capitalization of \$5 billion and above.

Table 14A reports results from the first stage of the 2SLS model above. Decimalization status is a significant instrument across all market capitalization (‘market cap’) categories. Across small, medium, and large cap stocks, decimalized stocks have 0.1-0.3% lower spreads. This result confirms the estimates obtained before. The sign on decimalization status for micro-cap stocks is positive and the spread is 6 bps higher for decimalized stocks. All first-stage F statistics are well over the value 10, the minimum F-value that gives us statistical significance of the excluded instrument. Table 14B reports results from the second stage of the 2SLS model. The coefficient on illiquidity is significant only in the case of large-cap stocks with a value of 0.5079. For a one standard deviation increase in illiquidity (σ=0.0078), next-day returns increase by 0.4% (=0.5079*0.0078). This is equivalent to an annual effect of 0.99% per year. Tables 15A and 15B estimates are the same as those in Tables 14A and B, except the standard errors are now clustered by stock. The results are qualitatively the same.

The weekly regressions with additional control variables from the HKV (2010) model are repeated under the scenario where Fama-French factors are directly included in the 2SLS model and the dependent variable is now weekly stock returns instead of weekly excess returns. The 2SLS model in equations 3a and 3b are modified as follows, where index *i* refers to the actual stock while index *j* refers to the 3-digit industry code:

$$E_t(R_{i,w+1}) = \alpha_i + \beta\, ESPR_{iw} + \omega_j^{MKTRF} MKTRF_w + \omega_j^{SMB} SMB_w + \omega_j^{HML} HML_w + \sum_{k=1}^{2} \gamma_{1k}\, R_{i,w-k} + \sum_{k=1}^{2} \gamma_{2k}\, R_{i,w-k}^{MKT} + \sum_{k=0}^{1} \gamma_{3k}\, \sigma_{i,w-k} + \sum_{k=0}^{1} \gamma_{4k}\, \sigma_{w-k}^{MKT} + \gamma_5 L_w^{MKT} + \epsilon_{i,w+1}, \qquad (5a)$$

$$ESPR_{iw} = \delta_i + \psi_j^{MKTRF} MKTRF_w + \psi_j^{SMB} SMB_w + \psi_j^{HML} HML_w + \sum_{k=1}^{2} \pi_{1k}\, R_{i,w-k} + \sum_{k=1}^{2} \pi_{2k}\, R_{i,w-k}^{MKT} + \sum_{k=0}^{1} \pi_{3k}\, \sigma_{i,w-k} + \sum_{k=0}^{1} \pi_{4k}\, \sigma_{w-k}^{MKT} + \pi_5 L_w^{MKT} + \mu Z_{iw} + \nu_{iw}, \qquad (5b)$$

Weekly stock volatility σ_{iw} and market volatility σ_w^{MKT} are estimated as outlined in French, Schwert, and Stambaugh (1987) and weekly estimates of other variables are obtained by averaging over the daily data. Tables 16A and 16B report results from the weekly 2SLS regression models. The instrument is a significant predictor of stock illiquidity in the first stage, and second stage results confirm that a one-standard deviation increase in stock illiquidity (σ=0.0288) increases stock returns by 1.2% in annual terms.[28]

2.5 Conclusion

The main contribution of this paper is to account for the endogeneity of liquidity in a standard asset pricing model. Since factors determining stock illiquidity such as

---

[28] It is interesting to note that weekly regressions on the entire sample with clustered standard errors (clustered at the firm-level) produced standard error estimates that rendered these coefficients statistically insignificant. Weekly regressions with firm-level clustered standard errors run on micro-, small-, mid-, and large-cap stocks revealed that these estimates are statistically significant for mid- and large-cap stocks.

macroeconomic factors or past stock market performance can also affect stock returns, we propose an instrumental variable based on the decimalization of NYSE, NASDAQ, and AMEX stocks over the 2000 to 2001 period. This instrument helps identify the effect of stock illiquidity on stock returns in the second stage and is a significant predictor of asset liquidity in the first stage. On average, decimalization decreased effective spreads by an estimated 20% and we found a 0.1% to 0.7% difference between the spread of a decimalized stock to that of a non-decimalized stock across different specifications. In the second stage of a 2SLS framework, a one standard deviation increase in stock illiquidity increases returns by 1.2% annually. With standard OLS regressions, we only get 7 to 9 basis points per year effect of illiquidity on returns. We find that modeling endogeneity of liquidity increases its effect on asset returns and from the OLS model, we can conclude that failing to account for this endogeneity can seriously underestimate the risk of illiquidity in a standard asset pricing model.

Liquidity can carry a very important role in valuing assets, and its role in the recent financial crisis has led to a renewed interest in understanding of liquidity in asset markets. The instrumental variable proposed in this paper can be useful in separately identifying the causal effect of stock liquidity on returns in any asset pricing model with liquidity. While this paper focused on pooled estimates, future work will consider the differential stock or industrial sector effect of liquidity. Furthermore, the coefficient on liquidity itself maybe time-varying in a returns equation (e.g., Watanabe and Watanabe (2008)) and modeling this time-varying causal effect of liquidity on asset returns would lead to a deeper understanding of low and high liquidity states in financial markets. Another fruitful extension of this paper would be to model liquidity and asset prices or returns in a dynamic sense, such as the vector-auto regression models in Pastor and Stambaugh (2009).

Table 2.1a: Decimalization Phases

NYSE (1053 firms in sample)		
Date	**No. of Stocks**	**Total**
28-Aug-00	7	7
25-Sep-00	57	64
4-Dec-00	95	159
29-Jan-01	All	-

NASDAQ (773 firms in sample)		
Date	**No. of Stocks**	**Total**
12-Mar-01	15	15
26-Mar-01	199	214
9-Apr-01	All	-

Note: NYSE and Amex had the same decimalization schedule. Only the decimalization phases for the NYSE and NASDAQ exchanges are reported here since Amex decimalization included many of the stocks listed on either the NYSE or the NASDAQ. Total number of stocks used in this sample is 2084. Approximately 87% (1826 stocks) of the stocks in the sample are listed on the NYSE or the NASDAQ.

Table 2.1b: Table of Means for Early and Late Decimalized Stocks

NYSE Stocks (Aug 1999 – Aug 2000)						
	Early (before Jan 29)		Late (on Jan 29)			
	No. of stocks = 84		No. of stocks = 601			
	No. of Obs = 19,714		No. of Obs = 138,411			
	Mean	**Std. Dev**	**Mean**	**Std. Dev**	**t-statistic**	**p-value**
$\boldsymbol{R_{it}}$ (%)	-0.054	3.152	-0.069	3.020	0.64	0.52
$\boldsymbol{R_{it}^{E}}$ (%)	-0.025	3.081	-0.041	2.935	0.71	0.48
$\boldsymbol{ESPR_{it}}$ (%)	0.852	1.199	0.913	1.344	-6.01	<0.0001
$\boldsymbol{Vol_{it}}$ **(# shares)**	1,138,004	3,604,949	530,018	1,320,479	45.01	<0.0001
$\boldsymbol{Size_{it}}$ **($ bn)**	8.04	20.6	5.60	19.6	16.27	<0.0001
NASDAQ Stocks (Mar 2000 – Mar 2001)						
	Early (before Apr 09)		Late (on Apr 09)			
	No. of stocks = 180		No. of stocks = 347			
	No. of Obs = 35,495		No. of Obs = 65,647			
	Mean	**Std. Dev**	**Mean**	**Std. Dev**	**t-statistic**	**p-value**
$\boldsymbol{R_{it}}$ (%)	-0.437	6.908	-0.153	5.371	-7.25	<0.0001
$\boldsymbol{R_{it}^{E}}$ (%)	-0.115	6.498	-0.057	5.194	-1.56	0.12
$\boldsymbol{ESPR_{it}}$ (%)	1.990	2.962	2.444	2.956	-23.31	<0.0001
$\boldsymbol{Vol_{it}}$ **(# shares)**	1,614,424	5,338,650	445,537	3,440,170	42.19	<0.0001
$\boldsymbol{Size_{it}}$ **($ bn)**	5.31	21.7	2.24	23.0	20.68	<0.0001
NASDAQ Stocks (Mar 1999 – Mar 2001)						
	Early (before Apr 09)		Late (on Apr 09)			
	No. of stocks = 180		No. of stocks = 389			
	No. of Obs = 65,212		No. of Obs = 137,158			
	Mean	**Std. Dev**	**Mean**	**Std. Dev**	**t-statistic**	**p-value**
$\boldsymbol{R_{it}}$ (%)	-0.132	6.335	-0.012	5.105	-4.56	<0.0001
$\boldsymbol{R_{it}^{E}}$ (%)	0.009	6.052	0.0004	4.993	0.34	0.74
$\boldsymbol{ESPR_{it}}$ (%)	1.928	2.815	2.455	2.890	-38.69	<0.0001
$\boldsymbol{Vol_{it}}$ **(# shares)**	1,273,004	4,391,590	341,855	2,663,461	58.93	<0.0001
$\boldsymbol{Size_{it}}$ **($ bn)**	4.30	17.9	1.88	19.3	27.08	<0.0001

Note: $\boldsymbol{R_{it}}$ is the daily return for a stock *i* on day *t*. $\boldsymbol{R_{it}^{E}}$ is the daily excess return for a stock *i* on day *t* and is calculated as a residual from the regression $(R_{it} - R_{ft}) = (\alpha_i + \beta_i^{MKTRF} \times MKTRF_t + \beta_i^{SMB} \times SMB_t + \beta_i^{HML} \times HML_t) + R_{it}^{E}$. $\boldsymbol{ESPR_{it}}$ is the relative effective spread of a stock, defined as twice the absolute difference between the transaction price and the bid-offer midpoint divided by the bid-offer midpoint. $\boldsymbol{Vol_{it}}$ is the number of shares traded in a day and $\boldsymbol{Size_{it}}$ is the market capitalization obtained by the product of price and number of shares outstanding.

Table 2.1c: Table of Means for Early and Late Decimalized Stocks - Demeaned Variables

	NYSE Stocks (Aug 1999 - Aug 2000)					
	Early (before Jan 29)		Late (on Jan 29)			
	No. of stocks = 84		No. of stocks = 601			
	No. of Obs = 19714		No. of Obs = 138,412			
	Mean	**Std. Dev**	**Mean**	**Std. Dev**	**t-statistic**	**p-value**
$\widetilde{R}_{it}$ (%)	-4.33E-10	3.147	3.76E-06	3.0123	-0.0002	0.99
$\widetilde{R}^E_{it}$ (%)	-0.025	3.0811	-0.041	2.9349	0.69	0.48
$\widetilde{ESPR}_{it}$ (%)	1.35E-09	0.8486	-4.85E-06	0.8959	0.0007	0.99
$\widetilde{Vol}_{it}$ (# shares)	-0.00388	1,941,580	0.064568	714,592	-0.01	0.99
$\widetilde{Size}_{it}$ ($)	-9.52	3.55bn	-76.1	2.69bn	-0.11	0.99

	NASDAQ Stocks (Mar 2000 – Mar 2001)					
	Early (before Apr 09)		Late (on Apr 09)			
	No. of stocks = 180		No. of stocks = 347			
	No. of Obs = 35,495		No. of Obs = 65,647			
	Mean	**Std. Dev**	**Mean**	**Std. Dev**	**t-statistic**	**p-value**
$\widetilde{R}_{it}$ (%)	-2.79E-09	6.8865	7.69E-10	5.359	-0.01	0.99
$\widetilde{R}^E_{it}$ (%)	-0.115	6.4979	-0.057	5.1942	-1.56	0.11
$\widetilde{ESPR}_{it}$ (%)	3.92E-09	2.0138	-3.24E-09	2.156	0.01	0.99
$\widetilde{Vol}_{it}$ (# shares)	0.008717	3,258,642	-0.00533	1,937,235	0.01	0.99
$\widetilde{Size}_{it}$ ($)	-18.0174	4.68bn	47.21737	6.04bn	0.01	0.99

	NASDAQ Stocks (Mar 1999 - Mar 2000)					
	Early (before Apr 09)		Late (on Apr 09)			
	No. of stocks = 180		No. of stocks = 389			
	No. of Obs = 65,212		No. of Obs = 137,158			
	Mean	**Std. Dev**	**Mean**	**Std. Dev**	**t-statistic**	**p-value**
$\widetilde{R}_{it}$ (%)	-2.16E-09	6.3239	-5.26E-10	5.1005	-0.01	0.99
$\widetilde{R}^E_{it}$ (%)	0.0091	6.0522	0.0004	4.9932	0.34	0.73
$\widetilde{ESPR}_{it}$ (%)	-1.82E-09	1.9969	9.19E-07	2.1769	-0.01	0.99
$\widetilde{Vol}_{it}$ (# shares)	0.012333	2,958,588	-0.0047	1,606,053	0.01	0.99
$\widetilde{Size}_{it}$ ($)	-5.24892	8.26bn	58.40356	5.78bn	-0.01	0.99

Note: $\widetilde{X_{it}} = X_{it} - \overline{X_i}$ denotes the demeaned variable. R_{it} is the daily return for a stock *i* on day *t*. R^E_{it} is the daily excess return for a stock *i* on day *t* and is calculated as a residual from the regression $(R_{it} - R_{ft}) = (\alpha_i + \beta_i^{MKTRF} \times MKTRF_t + \beta_i^{SMB} \times SMB_t + \beta_i^{HML} \times HML_t) + R^E_{it}$.

Table 2.2: Summary Statistics

Number of Stocks	2084
Maximum # Days	3020 (Mean = 1450)
Sample Period	Jan 1 1997 - Dec 31 2008

	N	Mean	Std. Dev
$\boldsymbol{R_{it}^{E}}$ **(in %)**	3,040,433	0.000	3.65
$\boldsymbol{ESPR_{it}}$ **(in %)**	3,041,978	1.31	2.92
$\boldsymbol{Dec\ Status_{it}}$	3,041,978	0.605	0.488
Price **(in $)**	3,041,978	26.9	34.6
$\boldsymbol{R_{it}}$ **(in %)**	3,040,433	0.012	3.74
$\boldsymbol{R_{it}^{CRSP}}$**(in %)**	2,998,963	0.058	3.82
$\boldsymbol{R_{t}^{MKT}}$ **(in %)**	3,041,978	0.013	1.26

Note: Sample consists of daily data on 2084 U.S. common stocks from January 1997 through December 2008. ***Maximum # Days*** is the maximum number of days of data available for each stock. On average, stocks have 1450 days of data. $\boldsymbol{R_{it}^{E}}$ is the daily excess return for a stock *i* on day *t* and is calculated as a residual from the regression $(R_{it} - R_{ft}) = (\alpha_i + \beta_i^{MKTRF} \times MKTRF_t + \beta_i^{SMB} \times SMB_t + \beta_i^{HML} \times HML_t) + R_{it}^{E}$. $\boldsymbol{ESPR_{it}}$ is the relative effective spread of a stock, defined as twice the absolute difference between the transaction price and the bid-offer midpoint divided by the bid-offer midpoint and $\boldsymbol{Dec\ Status_{it}}$ is an indicator variable whose value is 1 if the stock has been decimalized. ***Price*** is the average price adjusted for splits, dividends, etc. $\boldsymbol{R_{it}}$ is the average daily return across all stocks calculated by the author from data and $\boldsymbol{R_{i}^{CRSP}}$ is the daily return for a stock obtained from WRDS's CRSP database, calculated based on daily closing prices.

Table 2.3: Standard Pooled OLS Regression

	R^E_{it+1} (A)	R^E_{it+1} (B)	R^E_{it+1} (C)	R^E_{it+1} (D)
$ESPR_{it}$	0.0131***	0.0131***	0.0131***	0.0128***
	[0.0012]	[0.0012]	[0.0012]	[0.0012]
ΔADS_t	0.0061***	0.0062***	0.0066***	0.0065***
	[0.0008]	[0.0008]	[0.0008]	[0.001]
L^{MKT}_t	0.0061***	0.0061***	0.0058***	0.0047***
	[0.0003]	[0.0003]	[0.0003]	[0.0003]
TIME EFFECTS				
Day of week		X	X	X
Month			X	X
Year				X
Observations	3,037,378	3,037,378	3,037,378	3,037,378
R-squared	0.0003	0.001	0.001	0.001

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1 The time effects used in different specifications are clearly marked.

Table 2.4a: First-Stage Results

	$ESPR_{it}$ (A)	$ESPR_{it}$ (B)	$ESPR_{it}$ (C)	$ESPR_{it}$ (D)
$Dec\ Status_{it}$	-0.0028***	-0.0028***	-0.0029***	-0.0096***
	[0.00003]	[0.00003]	[0.00003]	[0.0002]
ΔADS_t	0.0059***	0.0058***	0.0047***	0.0044***
	[0.0006]	[0.0006]	[0.0006]	[0.0006]
L_t^{MKT}	-0.0076***	-0.0076***	-0.0082***	-0.0020***
	[0.0002]	[0.0002]	[0.0002]	[0.0002]
TIME EFFECTS				
Day of week		X	X	X
Month			X	X
Year				X
Observations	3,040,864	3,040,864	3,040,864	3,040,864
R-squared	0.0028	0.0028	0.0034	0.0149
Test of Excluded Instrument				
F(1,3040860)	8195.64	8196.22	8282.76	3636.73
p-value	<0.0001	<0.0001	<0.0001	<0.0001

Note: Robust standard errors are in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The time effects used in different specifications are clearly marked.

Table 2.4b: Second-Stage Results				
	R^{E}_{it+1} (A)	R^{E}_{it+1} (B)	R^{E}_{it+1} (C)	R^{E}_{it+1} (D)
$ESPR_{it}$	0.039**	0.039***	0.033**	0.058**
	[0.015]	[0.015]	[0.015]	[0.028]
ΔADS_t	0.0059***	0.0061***	0.0065***	0.0063***
	[0.001]	[0.001]	[0.001]	[0.001]
L^{MKT}_t	0.0063***	0.0063***	0.0061***	0.0048***
	[0.0003]	[0.0003]	[0.0003]	[0.0003]
TIME EFFECTS				
Day of week		X	X	X
Month			X	X
Year				X
Observations	3,037,378	3,037,378	3,037,378	3,037,378
Endogeneity Test				
Test statistic	2.930	3.097	1.884	2.569
p-value	0.0869	0.0784	0.1699	0.1090

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. The time effects used in different specifications are clearly marked.

Table 2.5a: First-Stage Estimates – Demeaned Variables

	$\widetilde{ESPR}_{\iota t}$ (A)	$\widetilde{ESPR}_{\iota t}$ (B)	$\widetilde{ESPR}_{\iota t}$ (C)	$\widetilde{ESPR}_{\iota t}$ (D)
$\widetilde{Dec\ Status}_{\iota}$	-0.0007***	-0.0007***	-0.0007***	-0.0012***
	[0.00003]	[0.00003]	[0.00003]	[0.00005]
$\Delta\widetilde{ADS}_t$	0.0063***	0.0063***	0.0052***	0.0041***
	[0.0005]	[0.0005]	[0.0005]	[0.0005]
$\widetilde{L_t^{MKT}}$	-0.0083***	-0.0083***	-0.0089***	-0.0035***
	[0.0002]	[0.0002]	[0.0002]	[0.0002]
TIME EFFECTS				
Day of week		X	X	X
Month			X	X
Year				X
Observations	3,037,378	3,037,378	3,037,378	3,037,378
R-squared	0.001	0.001	0.0018	0.0138
Test of excluded instrument				
F(1,3040860)	678.19	678.19	672.16	456.69
p-value	<0.0001	<0.0001	<0.0001	<0.0001

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. $\widetilde{X_{\iota t}} = X_{it} - \overline{X_{\iota}}$ denotes the demeaned variable. The time effects used in different specifications are clearly marked.

Table 2.5b: Second-Stage Estimates – Demeaned Variables				
	R^{E}_{it+1} (A)	R^{E}_{it+1} (B)	R^{E}_{it+1} (C)	R^{E}_{it+1} (D)
$\widetilde{ESPR_{\iota t}}$	0.2204***	0.2214***	0.2007***	0.2997***
	[0.069]	[0.069]	[0.070]	[0.087]
$\Delta\widetilde{ADS_t}$	0.0048***	0.0049***	0.0056***	0.0054***
	[0.0009]	[0.0009]	[0.0009]	[0.0009]
$\widetilde{L_t^{MKT}}$	0.0079***	0.0079***	0.0076***	0.0058***
	[0.0006]	[0.0006]	[0.0007]	[0.0005]
TIME EFFECTS				
Day of week		X	X	X
Month			X	X
Year				X
Observations	3,037,378	3,037,378	3,037,378	3,037,378
Endogeneity Test				
Test Statistic	8.384	8.474	6.792	10.389
p-value	0.0038	0.0036	0.0092	0.0013

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. $\widetilde{X_{\iota t}} = X_{it} - \overline{X_\iota}$ denotes the demeaned variable. The time effects used in different specifications are clearly marked.

Table 2.6a: First-Stage Estimates of Variables in Levels Over 2- and 4-Year Samples

	Sample Period			
	Jan 2000 - Dec 2001		Jan 1999 - Dec 2002	
	$\boldsymbol{ESPR_{it}}$	$\boldsymbol{\widetilde{ESPR_{it}}}$	$\boldsymbol{ESPR_{it}}$	$\boldsymbol{\widetilde{ESPR_{it}}}$
	(A)	(B)	(C)	(D)
$\boldsymbol{Dec\ Status_{it}}$	-0.0072***	-0.0039***	-0.0032***	-0.0007***
	[0.0001]	[0.0001]	[0.0001]	[0.0001]
$\boldsymbol{\Delta ADS_t}$	-0.0039**	-0.0093***	-0.0128***	-0.015***
	[0.0017]	[0.0012]	[0.0011]	[0.0008]
$\boldsymbol{L_t^{MKT}}$	0.0012*	-0.0031***	-0.0059***	-0.0065***
	[0.0006]	[0.0005]	[0.0003]	[0.0003]
TIME EFFECTS				
Day of week	X	X	X	X
Month	X	X	X	X
Observations	538,625	538,625	1,075,910	1,075,910
R-squared	0.0193	0.0165	0.0067	0.0059
Test of excluded instrument				
F test value	7312.22	4269.58	3805.79	278.89
p-value	<0.0001	<0.0001	<0.0001	<0.0001

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. $\boldsymbol{\widetilde{X_{it}} = X_{it} - \overline{X_i}}$ denotes the demeaned variable. The time effect used in different specifications are clearly marked**.** Columns A and C are the results for variables in levels while Columns B and D show the results for demeaned variables. Both left- and right-hand-side variables are demeaned in columns B and D.

Table 2.6b: Second-Stage Estimates of Variables in Levels Over 2- and 4-Year Samples

	Sample Period			
	Jan 2000 - Dec 2001		Jan 1999 - Dec 2002	
	$\boldsymbol{R^E_{it+1}}$	$\boldsymbol{R^E_{it+1}}$	$\boldsymbol{R^E_{it+1}}$	R^E_{it+1}
	(A)	(B)	(C)	(D)
$\boldsymbol{\widetilde{ESPR}_{it}}$	0.1321***	0.1045***	0.0742***	0.0444*
	[0.0188]	[0.019]	[0.024]	[0.0232]
$\boldsymbol{\Delta\widetilde{ADS}_t}$	0.0098***	0.028***	0.0066***	0.0101***
	[0.0026]	[0.0029]	[0.002]	[0.0016]
$\boldsymbol{\widetilde{L^{MKT}_t}}$	0.0129***	0.0133***	0.009***	0.0063***
	[0.001]	[0.0010]	[0.0004]	[0.0005]
TIME EFFECTS				
Day of week	X	X	X	X
Month		X		X
Observations	538,625	538,625	1,075,910	1,075,910
Endogeneity Test				
Test Statistic	31.915	16.829	4.034	0.671
p-value	<0.0001	<0.0001	0.0446	0.4125

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.7: Second-Stage Estimates of Demeaned Variables Over 2- and 4-Year Samples

	Sample Period			
	Jan 2000 - Dec 2001		Jan 1999 - Dec 2002	
	R^{E}_{it+1}	R^{E}_{it+1}	R^{E}_{it+1}	R^{E}_{it+1}
	(A)	(B)	(C)	(D)
$\widetilde{ESPR}_{\iota t}$	0.2798***	0.2263***	0.4492***	0.2677**
	[0.036]	[0.035]	[0.129]	[0.113]
$\Delta\widetilde{ADS}_t$	0.0109***	0.0293***	0.0103***	0.0136***
	[0.0026]	[0.0029]	[0.0020]	[0.0024]
$\widetilde{L^{MKT}_t}$	0.0134***	0.0142***	0.0110***	0.0078***
	[0.001]	[0.0011]	[0.001]	[0.001]
TIME EFFECTS				
Day of week	X	X	X	X
Month		X		X
Observations	538,027	538,027	1,075,910	1,075,910
Endogeneity Test				
Test Statistic	34.049	18.907	10.121	3.873
p-value	<0.0001	<0.0001	0.0015	0.0491

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. $\widetilde{X_{\iota t}} = X_{it} - \overline{X_\iota}$ denotes the demeaned variable. The time effects used in different specifications are clearly marked**.** Columns A and C are the results for variables in levels while Columns B and D show the results for demeaned variables.

Table 2.8a: First-Sage Estimates from Regressions on Weekly Data

	$ESPR_{i,w}$
ΔADS_w	0.0005
	[0.0016]
$R_{i,w-1}$	-0.0575***
	[0.0038]
$R_{i,w-2}$	-0.0374***
	[0.0035]
R^{MKT}_{w-1}	0.0132***
	[0.0033]
R^{MKT}_{w-2}	0.0109*
	[0.0032]
Dec_{iw}	-0.0074***
	[0.0003]
L^{MKT}_{w}	-0.0004
	[0.0005]
$\sigma_{i,w}$	0.0634***
	[0.0014]
$\sigma_{i,w-1}$	0.0476***
	[0.0013]
σ^{MKT}_{w}	-0.0014
	[0.0033]
σ^{MKT}_{w-1}	0.0005
	[0.0034]
TIME EFFECTS	
Month	X
Year	X
Observations	659,294
R-squared	0.0628
Test of excluded instrument	
F test value	584.01
p-value	<0.0001

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.8b: Second-Stage Estimates rom Regressions on Weekly Data

	$R_{i,w+1}$
$ESPR_{i,w}$	-0.0071
	[0.040]
ΔADS_{w}	-0.0089***
	[0.0011]
$R_{i,w-1}$	-0.0104**
	[0.0044]
$R_{i,w-2}$	-0.0050
	[0.0038]
R_{w-1}^{MKT}	-0.0006
	[0.0025]
R_{w-2}^{MKT}	-0.0129***
	[0.0025]
L_{w}^{MKT}	0.0059***
	[0.0004]
$\sigma_{i,w}$	0.0057*
	[0.0032]
$\sigma_{i,w-1}$	-0.0048*
	[0.0025]
σ_{w}^{MKT}	-0.0099***
	[0.0029]
σ_{w-1}^{MKT}	-0.0006
	[0.0027]
TIME EFFECTS	
Month	X
Year	X
Observations	659,294
Endogeneity Test	
Test Statistic	0.020
p-value	0.8885

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.9a: First-Stage Estimates of Demeaned Variables from Regressions on Weekly Data

	$\widetilde{ESPR_{\iota,w}}$
$\Delta\widetilde{ADS_w}$	0.0001
	[0.0012]
$\widetilde{R_{\iota,w-1}}$	-0.0373***
	[0.0028]
$\widetilde{R_{\iota,w-2}}$	-0.0280***
	[0.0026]
$\widetilde{R_{w-1}^{MKT}}$	0.0335***
	[0.0057]
$\widetilde{R_{w-2}^{MKT}}$	0.0365***
	[0.0056]
$\widetilde{Dec_{\iota w}}$	-0.0008***
	[0.0001]
$\widetilde{L_w^{MKT}}$	-0.0013***
	[0.0003]
$\widetilde{\sigma_{\iota,w}}$	0.0314***
	[0.0008]
$\widetilde{\sigma_{\iota,w-1}}$	0.0127***
	[0.0008]
$\widetilde{\sigma_w^{MKT}}$	0.0384***
	[0.0025]
$\widetilde{\sigma_{w-1}^{MKT}}$	0.0425
	[0.0026]
Month	X
Year	X
Observations	659,294
R-squared	0.036
Test of excl. instrument	
F test value	68.46
p-value	<0.0001

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. $\widetilde{X_{\iota t}} = X_{it} - \overline{X_\iota}$ denotes the demeaned variable.

Table 2.9b: Second-Stage Estimates of Demeaned Variables from Regressions on Weekly Data

	$R_{i,w+1}$
$\widetilde{ESPR_{\iota,w}}$	0.8522***
	[0.2440]
$\Delta\widetilde{ADS_w}$	-0.0164***
	[0.0015]
$\widetilde{R_{\iota,w-1}}$	-0.0684
	[0.114]
$\widetilde{R_{\iota,w-2}}$	0.0107
	[0.0101]
$\widetilde{R_{w-1}^{MKT}}$	0.0158
	[0.0155]
$\widetilde{R_{w-2}^{MKT}}$	-0.0004
	[0.0100]
$\widetilde{L_{w}^{MKT}}$	0.0104***
	[0.0006]
$\widetilde{\sigma_{\iota,w}}$	-0.0186
	[0.0082]
$\widetilde{\sigma_{\iota,w-1}}$	-0.0148***
	[0.0036]
$\widetilde{\sigma_{w}^{MKT}}$	-0.02***
	[0.0091]
$\widetilde{\sigma_{w-1}^{MKT}}$	-0.0451***
	[0.0112]
TIME EFFECTS	
Month	X
Year	X
Observations	659,465

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1.$\widetilde{X_{\iota t}} = X_{it} - \overline{X_\iota}$ denotes the demeaned variable.

Table 2.10: 2SLS with Newey-West Standard Errors and Demeaned Variables on Daily Data

	R^E_{it+1} OLS	R^E_{it+1} (A)	R^E_{it+1} (B)	R^E_{it+1} (C)	R^E_{it+1} (D)
$\widetilde{ESPR}_{it}$	0.019***	0.220***	0.221***	0.201***	0.300***
	[0.001]	[0.063]	[0.063]	[0.063]	[0.083]
$\Delta\widetilde{ADS}_t$	0.006***	0.005***	0.005***	0.006***	0.005***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
$\widetilde{L^{MKT}_t}$	0.006***	0.008***	0.008***	0.008***	0.006***
	[0.000]	[0.001]	[0.001]	[0.001]	[0.000]
TIME EFFECTS					
Day of week	X		X	X	X
Month	X			X	X
Year	X				X
Observations	3,037,378	3,037,378	3,037,378	3,037,378	3,037,378

Note: Lag length for Newey-West standard errors is 11. Standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. $\widetilde{X_{it}} = X_{it} - \overline{X_i}$ denotes the demeaned variable. Columns A through D show the second stage results from a 2SLS regression. The time effects used in different specifications are clearly marked.

Table 2.11: 2SLS with Clustered Standard Errors and Demeaned Variables on Daily Data (Clustering at the Firm Level)

	R^E_{it+1} OLS	R^E_{it+1} (A)	R^E_{it+1} (B)	R^E_{it+1} (C)	R^E_{it+1} (D)
$\widetilde{ESPR}_{it}$	0.019***	0.220*	0.221*	0.201*	0.300***
	[0.001]	[0.126]	[0.126]	[0.118]	[0.100]
$\Delta\widetilde{ADS}_t$	0.006***	0.005***	0.005***	0.006***	0.005***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
$\widetilde{L_t^{MKT}}$	0.005***	0.008***	0.008***	0.008***	0.006***
	[0.0003]	[0.001]	[0.001]	[0.001]	[0.001]
TIME EFFECTS					
Day of week	X		X	X	X
Month	X			X	X
Year	X				X
Observations	3,037,378	3,037,378	3,037,378	3,037,378	3,037,378

Note: Standard errors clustered at the firm level are in brackets. *** p<0.01, ** p<0.05, * p<0.1. $\widetilde{X_{it}} = X_{it} - \overline{X_i}$ denotes the demeaned variable. Columns A through D show the second stage results from a 2SLS regression. The time effects used in different specifications are clearly marked.

Table 2.12: Results for NYSE Stocks

Panel A: Summary Statistics

Variable	Obs	Mean	Std. Dev.
$\boldsymbol{R_{it}}$	1,709,065	-0.00016	0.0264
$\boldsymbol{R_{it}^{E}}$	1,709,065	2.45E-07	0.0253
$\widetilde{\boldsymbol{ESPR}_{it}}$	1,709,090	-6.6E-05	0.0133

Panel B: Regression Results

	$\boldsymbol{R_{it+1}^{E}}$	$\boldsymbol{R_{it+1}^{E}}$	$\boldsymbol{R_{it+1}^{E}}$
	Robust SE	SE cluster by firm	Newey-West SE
$\widetilde{\boldsymbol{ESPR}_{it}}$	0.376***	0.376**	0.376***
	[0.117]	[0.153]	[0.125]
$\Delta\widetilde{\boldsymbol{ADS}_{t}}$	0.008***	0.008***	0.008***
	[0.001]	[0.001]	[0.001]
$\widetilde{\boldsymbol{L_{t}^{MKT}}}$	0.005***	0.005***	0.005***
	[0.000]	[0.000]	[0.000]
TIME EFFECTS			
Day of week	X	X	X
Month	X	X	X
Year	X	X	X
Observations	1,707,440	1,707,440	1,707,440

Note: Standard errors calculated via different methods are in brackets. *** p<0.01, ** p<0.05, * p<0.1. $\widetilde{\boldsymbol{X}_{it}} = \boldsymbol{X}_{it} - \overline{\boldsymbol{X}_{i}}$ denotes the demeaned variable. Panel A shows the relevant summary statistics. In Panel B, columns show the second stage results from a 2SLS regression with different standard error calculation specifications. The time effects used in different specifications are clearly marked**.**

Table 2.13: Results for NASDAQ Stocks

Panel A: Summary Statistics

Variable	Obs	Mean	Std. Dev.
$\boldsymbol{R_{it}}$	1,331,368	-0.0004	0.0479
$\boldsymbol{R_{it}^{E}}$	1,331,368	1.68E-06	0.0471
$\widetilde{\boldsymbol{ESPR}_{\iota t}}$	1,332,888	0.000084	0.0337

Panel B: Regression Results

VARIABLES	$\boldsymbol{R_{it+1}^{E}}$	$\boldsymbol{R_{it+1}^{E}}$	$\boldsymbol{R_{it+1}^{E}}$
	Robust SE	SE cluster by firm	Newey-West SE
$\widetilde{\boldsymbol{ESPR}_{\iota t}}$	0.2065**	0.2065**	0.2065**
	[0.1044]	[0.1063]	[0.099]
$\Delta\widetilde{\boldsymbol{ADS}_{t}}$	0.0027	0.0027	0.0027
	[0.0018]	[0.0018]	[0.0018]
$\widetilde{\boldsymbol{L_{t}^{MKT}}}$	0.0068***	0.0068***	0.0068***
	[0.001]	[0.001]	[0.001]
TIME EFFECTS			
Day of week	X	X	X
Month	X	X	X
Year	X	X	X
Observations	989,418	989,418	989,418

Note: Standard errors calculated via different methods are in brackets. *** p<0.01, ** p<0.05, * p<0.1. $\widetilde{\boldsymbol{X}_{\iota t}} = \boldsymbol{X}_{it} - \overline{\boldsymbol{X}_{\iota}}$ denotes the demeaned variable. Panel A shows the relevant summary statistics. In Panel B, columns show the second stage results from a 2SLS regression with different standard error specifications. The time effects used in different specifications are clearly marked**.**

Table 2.14a: First-Stage Estimates of Demeaned Variables from Regressions on Daily Data (By Market Cap)

	MARKET CAP			
	Micro	**Small**	**Mid**	**Large**
	$ESPR_{it}$	$ESPR_{it}$	$ESPR_{it}$	$ESPR_{it}$
$Dec\ Status_i$	0.0006***	-.0028***	-.0018***	-.0010***
	[0.0001]	[.0001]	[.0001]	[.0001]
ΔADS_t	0.0075***	0.0047***	0.0022***	0.0011***
	[.0015]	[0.0008]	[.0004]	[.0003]
L_t^{MKT}	-.0082***	-.0025***	-.0006***	-.0010***
	[.0005]	[0.0002]	[.0001]	[0.0001]
TIME EFFECTS				
Day of Week	X	X	X	X
Month	X	X	X	X
Year	X	X	X	X
Number of stocks	921	490	424	258
Total No. of Observations	989,233	818,226	749,749	486,799
Test of excluded instrument				
F test	24.75	1192.44	1697.48	590.89
p-value	<0.0001	<0.0001	<0.0001	<0.0001

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. Each column represents a single regression run on stocks sorted in to micro-cap (market capitalization < $250 million), small-cap (between $250 million and $2 billion), mid-cap ($2-$5 billion), and large-cap ($5 billion and over.) Time fixed effects and coefficients on Fama-French factors have been suppressed. Variables have been demeaned by subtracting out the panel level mean wherever appropriate.

Table 2.14b: Second-Stage Estimates of Demeaned Variables from Regressions on Daily Data (By Market Cap)

	MARKET CAP			
	Micro	**Small**	**Mid**	**Large**
	R_{it+1}	R_{it+1}	R_{it+1}	R_{it+1}
$ESPR_{it}$	-0.6577	0.0520	0.1054	0.5079*
	[0.3405]	[0.0669]	[0.0959]	[0.2725]
ΔADS_t	0.0010	-.0045***	-.0033**	-0.0027
	[0.0031]	[0.0016]	[0.0014]	[0.0017]
L_t^{MKT}	0.0003	0.0028***	0.0033***	0.0039***
	[0.0028]	[0.0007]	[0.0006]	[0.0008]
TIME EFFECTS				
Day of Week	X	X	X	X
Month	X	X	X	X
Year	X	X	X	X
Number of stocks	921	490	424	258
Total No. of Observations	988,011	817,709	749,281	482,377

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. Each column represents a single regression run on stocks sorted in to micro-cap (market capitalization < $250 million), small-cap (between $250 million and $2 billion), mid-cap ($2-$5 billion), and large-cap ($5 billion and over.) Time fixed effects and coefficients on Fama-French factors have been suppressed. Variables have been demeaned by subtracting out the panel level mean wherever appropriate.

Table 2.15a: First-Stage Estimates of Demeaned Variables from Regressions on Daily Data (By Market Cap)

	MARKET CAP			
	Micro	**Small**	**Mid**	**Large**
	$ESPR_{it}$	$ESPR_{it}$	$ESPR_{it}$	$ESPR_{it}$
$Dec\ Status_i$	0.0006	-.0028***	-.0018***	-.0010***
	[0.0009]	[.0005]	[.0003]	[.0001]
ΔADS_t	0.0075***	0.0047***	0.0022***	0.0011**
	[.0020]	[0.0009]	[.0006]	[.0004]
L_t^{MKT}	-.0082***	-.0025***	-.0006***	-.0010***
	[.0008]	[0.0004]	[.0003]	[0.0003]
TIME EFFECTS				
Day of Week	X	X	X	X
Month	X	X	X	X
Year	X	X	X	X
Number of stocks	921	490	424	258
Total No. of Observations	989,233	818,226	749,749	486,799
Test of excluded instrument				
F test	0.47	24.54	41.42	37.42
p-value	0.4935	<0.0001	<0.0001	<0.0001

Note: Standard errors clustered at the firm-level are in brackets. *** p<0.01, ** p<0.05, * p<0.1. Each column represents a single regression run on stocks sorted in to micro-cap (market capitalization < $250 million), small-cap (between $250 million and $2 billion), mid-cap ($2-$5 billion), and large-cap ($5 billion and over.) Time fixed effects and coefficients on Fama-French factors have been suppressed. Variables have been demeaned by subtracting out the panel level mean wherever appropriate.

Table 2.15b: Second-Stage Estimats of Demeaned Variables from Regressions on Daily Data (By Market Cap)

	MARKET CAP			
	Micro	**Small**	**Mid**	**Large**
	R_{it+1}	R_{it+1}	R_{it+1}	R_{it+1}
$ESPR_{it}$	-.6577	0.0520	0.1054*	0.5079***
	[0.9948]	[0.0406]	[0.0576]	[0.1716]
ΔADS_t	0.0010	-.0045***	-.0033***	-0.0027**
	[0.0072]	[0.0011]	[0.0012]	[0.0012]
L_t^{MKT}	0.0003	0.0028***	0.0033***	0.0039***
	[0.0079]	[0.0005]	[0.0005]	[0.0006]
TIME EFFECTS				
Day of Week	X	X	X	X
Month	X	X	X	X
Year	X	X	X	X
Number of stocks	921	490	424	258
Total No. of Observations	988,011	817,709	749,281	482,377

Note: Standard errors clustered at the firm level are in brackets. *** p<0.01, ** p<0.05, * p<0.1. Each column represents a single regression run on stocks sorted in to micro-cap (market capitalization < $250 million), small-cap (between $250 million and $2 billion), mid-cap ($2-$5 billion), and large-cap ($5 billion and over.) Time fixed effects and coefficients on Fama-French factors have been suppressed. Variables have been demeaned by subtracting out the panel level mean wherever appropriate.

Table 2.16a: First-Stage Estimates of Demeaned Variables from Regressions on Weekly Data (One-Step)

	$\widetilde{ESPR_{\iota,w}}$
$\Delta\widetilde{\boldsymbol{ADS_w}}$	0.0001
	[.0012]
$\widetilde{\boldsymbol{R_{\iota,w-1}}}$	-0.0418***
	[.0031]
$\widetilde{\boldsymbol{R_{\iota,w-2}}}$	-0.0385***
	[.0027]
$\widetilde{\boldsymbol{R_{w-1}^{MKT}}}$	0.0551***
	[.0068]
$\widetilde{\boldsymbol{R_{w-2}^{MKT}}}$	0.0256***
	[0.0056]
$\widetilde{\boldsymbol{Dec_{\iota w}}}$	-0.0005***
	[.0001]
$\widetilde{\boldsymbol{L_{w}^{MKT}}}$	-0.0014***
	[.0004]
$\widetilde{\boldsymbol{\sigma_{\iota,w}}}$	0.0319***
	[.0008]
$\widetilde{\boldsymbol{\sigma_{\iota,w-1}}}$	0.0136***
	[.0008]
$\widetilde{\boldsymbol{\sigma_{w}^{MKT}}}$	0.0343***
	[.0028]
$\widetilde{\boldsymbol{\sigma_{w-1}^{MKT}}}$	0.0438***
	[.0027]
TIME EFFECTS	
Month	X
Year	X
Observations	661,370
R-squared	0.036
Test of excluded instrument	
F test value	29.16
p-value	<0.0001

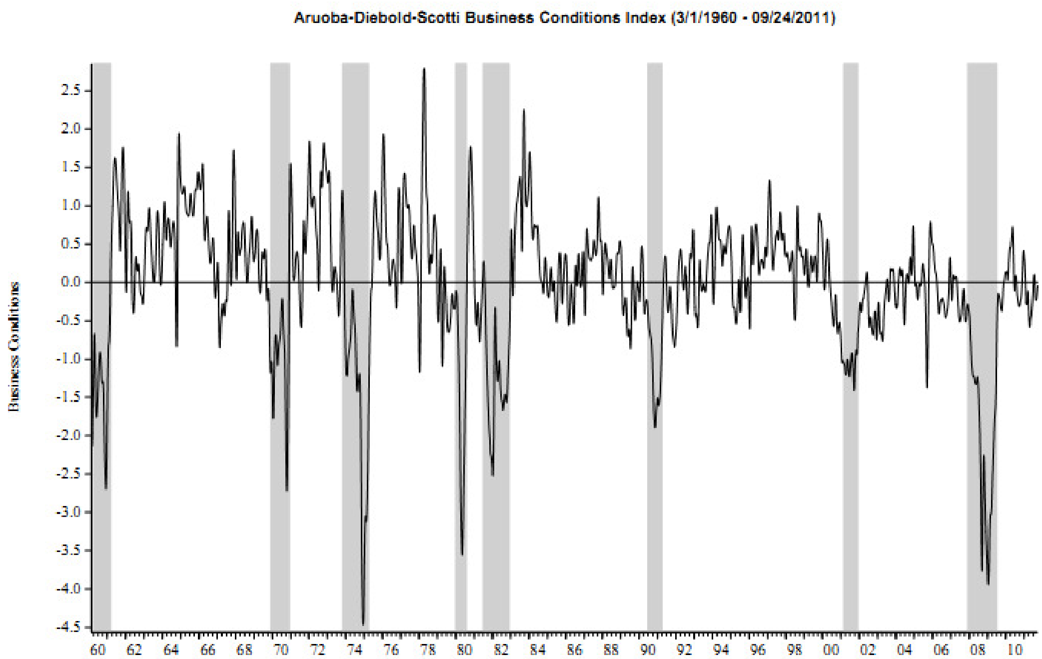
Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. Time fixed effects and coefficients on Fama-French factors have been suppressed. Variables have been demeaned by subtracting out the panel level mean wherever appropriate.

Table 2.16b: Second-Stage Estimates of Demeaned Variables from Regressions on Weekly Data (One-Step)

	$R_{i,w+1}$
$\widetilde{ESPR_{\iota,w}}$	0.8415***
	[0.2416]
$\Delta\widetilde{ADS_w}$	-0.0164***
	[.0015]
$\widetilde{R_{\iota,w-1}}$	-0.0690***
	[0.0113]
$\widetilde{R_{\iota,w-2}}$	0.0104
	[0.0100]
$\widetilde{R_{w-1}^{MKT}}$	0.0163
	[0.0154]
$\widetilde{R_{w-2}^{MKT}}$	-0.0003
	[.0100]
$\widetilde{L_w^{MKT}}$	0.0103***
	[0.0006]
$\widetilde{\sigma_{\iota,w}}$	-0.0183**
	[0.0081]
$\widetilde{\sigma_{\iota,w-1}}$	-0.0147***
	[.0036]
$\widetilde{\sigma_w^{MKT}}$	-0.0196**
	[.0090]
$\widetilde{\sigma_{w-1}^{MKT}}$	-0.0446***
	[.0111]
TIME EFFECTS	
Month	X
Year	X
Observations	658,779

Note: Robust standard errors are in brackets. *** p<0.01, ** p<0.05, * p<0.1. Time fixed effects and coefficients on Fama-French factors have been suppressed. Variables have been demeaned by subtracting out the panel level mean wherever appropriate.

2.6 Appendix



Note: We construct the ADS Index using the latest data available as of September 30, 2011. This includes (1) initial jobless claims through the week ending September 24, 2011, (2) payroll employment through August 2011, (3) industrial production through August 2011, (4) real personal income through August 2011, (5) real manufacturing and trade sales through July 2011, and (6) real GDP through the second quarter of 2011. Gray shading indicates NBER-designated recessions. The limits used on the y axis reflect the minimum and maximum values of the index over its entire history.

Figure 2.A1: Aruoba-Diebold-Scotti Business Index Measure

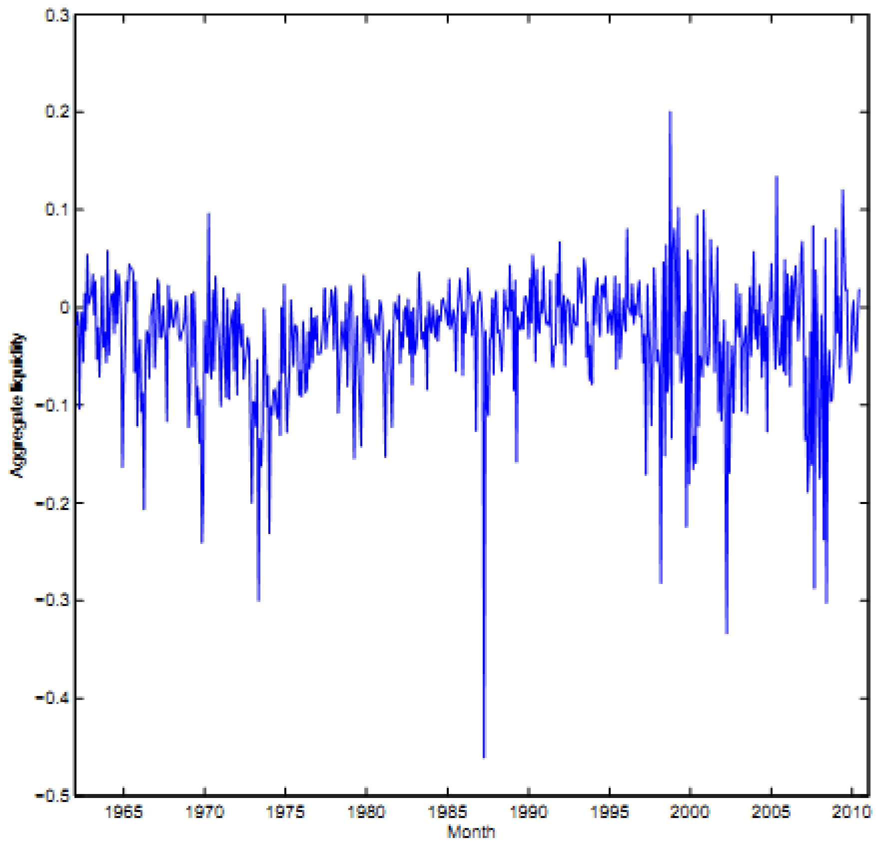


Figure 2.A2: Pastor-Stambaugh Aggregate Liquidity Measure

Note: Pastor and Stambaugh build their liquidity factor by estimating for each stock i and in each month t, the regression of daily excess stock return over the market on a constant, the previous day return and the dollar volume in the previous day multiplied by the sign of the previous day return. The regression coefficient of the return on the signed volume is the measure of the individual stock liquidity. The market wide liquidity is simply the cross-sectional average of the individual stock liquidity coefficients. The market liquidity factor is constructed by weighting the market liquidity with the ratio of total dollar value of the market in the previous month and total dollar market value at the beginning of the sample. Innovations in this aggregate liquidity factor is plotted here.

2.7 References

Acharya, Viral V., and Lasse Heje Pedersen, 2005, "Asset Pricing with Liquidity Risk," *Journal of Financial Economics* 77(2), pp. 375-410.

Amihud, Yakov, and Haim Mendelson, 1986, "Asset pricing and the bid-ask spread," *Journal of Financial Economics* 17(2), pp. 223-249.

Anshuman, Ravi, and S. Viswanathan, 2005, "Market Liquidity," Working Paper, Duke University.

Aruoba, S. Boragan, Francis X. Diebold, and Chiara Scotti, 2009, "Real-Time Measurement of Business Conditions," *Journal of Business and Economic Statistics* 27(4), pp. 417-427.

Baltagi. Badi Hani, *Nonstationary panels, panel cointegration, and dynamic panels*. New York: Elsevier, 2000.

Bencivenga, Valerie R., Bruce D. Smith, and Ross M. Starr, 1995, "Transactions Costs, Technological Choice, and Endogenous Growth," *Journal of Economic Theory* 67(1), pp. 153-177.

Bencivenga, Valerie R., Bruce D. Smith, and Ross M. Starr, 1996, "Equity Markets, Transactions Costs, and Capital Accumulation: An Illustration," *World Bank Economic Review* 10(2), pp. 241-265.

Bessembinder, Hendrik, 2003, "Trade Execution Costs and Market Quality after Decimalization," *Journal of Financial and Quantitative Analysis* 38(4), pp. 747-777.

Brock, W., and Allan Kleidon, 1992, "Periodic market closure and trading volume: A model of intraday bid and asks," *Journal of Economic Dynamics and Control* 16(3-4), pp. 451-489.

Brunnermeier, Markus K., and Lasse Heje Pedersen, 2009, "Market Liquidity and Funding Liquidity," *Review of Financial Studies* 22(6), pp. 2201-2238.

Cenesizoglu, Tolga, 2011, "Size, book-to-market ratio and macroeconomic news," *Journal of Empirical Finance* 18(2), pp. 248-270.

Chakravarty, Sugato, Stephen P. Harris, and Robert A. Wood, 2001, "Decimal Trading and Market Impact," Working Paper, Purdue University.

Chan, K.C., William G. Christie, and Paul H. Schultz, 1995, "Market Structure and the Intraday Pattern of Bid-Ask Spreads for NASDAQ Securities," *The Journal of Business* 68(1), pp. 35-60.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, “Commonality in Liquidity,” *Journal of Financial Economics* 56(1), pp. 3-28.

Chordia, Tarun, Asani Sarkar, and Avanidhar Subrahmanyam, 2005, “An Empirical Analysis of Stock and Bond Market Liquidity,” *Review of Financial Studies* 18, pp.85-129.

Eisfeldt, Andrea, 2004, “Endogenous Liquidity in Asset Markets,” *Journal of Finance* 59(1), pp. 1-30.

Fama, Eugene F.; French, Kenneth R. (1993). "Common Risk Factors in the Returns on Stocks and Bonds". *Journal of Financial Economics* 33 (1), pp. 3–56 .

French, Kenneth R., William G. Schwert, and Robert Stambaugh, 1987, "Expected Stock Returns and Volatility," *Journal of Financial Economics* 19(1), pp. 3-29.

Garleanu, Nicolae, and Lasse Heje Pedersen, 2007, "Liquidity and risk management," *American Economic Review* 97(2), pp. 193–197.

Goldstein, Michael A., and Kenneth A. Kavajecz, 2000, "Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE," *Journal of Financial Economics* 56(1), pp. 125-149.

Gromba, Denis, and Dimitri Vayanos, 2002, "Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs," *Journal of Financial Economics* 66(2-3), pp. 361-407.

Hamilton, James D., and Gang Lin, 1996, "Stock Market Volatility and the Business Cycle," *Journal of Applied Econometrics* 11(5), pp. 573–593.

Hamilton, James D., 2009, “The Aruoba-Diebold-Scotti Business Index” *Econbrowser*, http://www.econbrowser.com/

Hamilton, James D., 2011, “Real-time analysis of the Aruoba-Diebold-Scotti Business Conditions Index” *Econbrowser*, http://www.econbrowser.com/

Hameed, Allaudeen, Wenjin Kang, and S. Viswanathan, 2010, "Stock Market Declines and Liquidity," *Journal of Finance* 65(1), pp. 257-293.

Hasbrouck, Joel, and Duane J. Seppi, 2001, “Common factors in prices, order flows, and liquidity,” *Journal of Financial Economics* 59, pp.383-411.

Huberman, Gur, and Dominika Halka, 2001, “Systematic liquidity,” *Journal of Financial Research* 24, pp. 161-178.

Korajczyka, Robert A., and Ronnie Sadka, 2008, "Pricing the commonality across alternative measures of liquidity," *Journal of Financial Economics* 87(1), pp. 45-72.

Kyle, Albert S., and Wei Xiong, 2001, "Contagion as a Wealth Effect," *Journal of Finance* 56(4), pp. 1401-1440.

Mitchell, Mark, Lasse Heje Pedersen, and Todd Pulvino, 2007, "Slow moving capital," *American Economic Review* 97(2), pp. 215–220.

Norris, Floyd, "So Long, Fractions, but Maybe Not Till 2000," *The New York Times*, June 6, 1997.

Norris, Floyd, "May Day Redux: Will Decimalization decimate Profits?," *The New York Times*, August 25, 2000.

Pastor, Lubos, and Robert F. Stambaugh, 2003, "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy* 111(3), pp. 642-685.

Pastor, Lubos, and Robert F. Stambaugh, 2009, "Predictive Systems: Living with Imperfect Predictors," *Journal of Finance* 64(4), pp. 1583-1628.

Petersen, Mitchell A., 2009, "Estimating Standard Errors in Finance Panel Datasets: Comparing Apporaches," *Review of Financial Studies* 22, pp. 453-480.

Stoll, Hans R., 1978, "The Supply of Dealer Services in Securities Markets," *Journal of Finance* 33(4), pp. 1133-1151.

Vayanos, Dimitri, 2004, "Flight To Quality, Flight to Liquidity and the Pricing of Risk," NBER Working Paper No. w10327.

Watanabe, Akiko and Masahiro Watanabe, 2008, "Time-Varying Liquidity Risk and the Cross Section of Stock Returns," *Review of Financial Studies* 21(6), pp. 2449-2486.

White, Halbert, 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48(4), pp.817-838.

Wood, Robert, Thomas H. McInish, and J.Keith Ord, 1985, "An Investigation of Transaction Data for NYSE Stocks," *Journal of Finance* 40(3), pp. 723-739.

3. Measuring Algorithmic Trading

3.1 Introduction

Spurred by technological innovations, there have been significant shifts in the structure of stock market trading in recent years. For example, the use of automated computer programs to monitor stock prices and execute trades, known as algorithmic trading (AT), has become much more prevalent.[29] This type of trading has become notorious due to one particular kind, high frequency trading, where stock positions turn over rapidly—often within seconds or hours. The recent financial crisis and many small market crashes, including the "Flash Crash" of May 2010, have led to a debate over whether such trading mechanisms tend to stabilize the market by providing additional liquidity during times of need, or destabilize the market by increasing price volatility. If algorithmic and high frequency trading tend to destabilize the market, then taxing certain types of financial transaction may lead to more efficient markets by keeping speculative trading in check. But if these transactions tend to provide liquidity during times of need, then taxation would actually hamper financial market efficiency by reducing market liquidity when it is needed most.

Despite the importance for financial market regulation, very little is known empirically about the effects of algorithmic or high frequency trading on market quality. This paper will study the effect of AT in particular on stock liquidity in an attempt to shed some light on this question. In general, it is difficult to identify whether a particular trade was undertaken by a computer or a human trader. In order to proxy for AT, we propose two related measures constructed using data from the SEC-mandated Disclosure of Order Execution (DOE) dataset: the average time taken to execute an order for a particular stock on the NYSE,

[29] An article in the *Financial Times* (July 29, 2009) reports that high-frequency trading, a subset of algorithmic trading, accounts for over 73% of trading volume in the U.S. equity markets.

and the proportion of orders filled within ten seconds of order receipt. To the best of our knowledge, these measures have not been previously used as proxies for AT, though they have been used to compare order execution qualities across various U.S market centers.[30] We argue that these measures can be a good proxy for AT as the DOE dataset is put together from actual orders received and executed in the NYSE, and we are able to identify marketable and non-marketable orders (Korajczyk and Sadka (2008).)[31]

Using these proxies and data from the NYSE, we estimate the impact of AT on stock spread using several identification approaches. First, we use a fixed effects model to control for unobservable heterogeneity across stocks. However, the AT measure may be related to unobserved determinants of liquidity—for example, time to fill may itself be a function of the stock's spread— and so be correlated with unobserved determinants of the spread. . To overcome this endogeneity problem, the second identification approach follows Hendershott, Jones, and Menkveld (2011) in instrumenting for the measure of AT using the NYSE's introduction of automatically disseminated liquidity quotes (known as "autoquote"). First stage results confirm the relevance of this instrument. At the monthly level, the move from manual to automated quotes decreased average order fill times by 5 seconds per order. Accounting for the endogeneity of AT, we find that a 1 second decrease in average time taken to fill an order narrows spreads by 2 basis points (0.02 %.) The OLS estimate is many orders of magnitude smaller at 0.0002%. The difference in the OLS and 2SLS estimates points to the negative correlation between AT and an unobservable determinant of stock illiquidity, and is further discussed in the results section.

---

[30] See Boehmer (2005), Bessembinder (2003), He et al (2006), among others.

[31] Marketable orders are either market orders with an order price at the current market price, or limit orders with an order price set above the current price for a buy order and below the current price for a sell order. Nonmarketable orders are buy (sell) limit orders with an order price below (above) the market. Marketable orders remove liquidity from the market while nonmarketable orders add liquidity to the market.

Though the literature on the topic of AT is relatively new, several papers yield some interesting results. Hendershott, Jones, and Menkveld (2011) propose the number of electronic messages per $100 of trading volume as a proxy for algorithmic trading and instrument for endogeneity in this measure using NYSE's autoquote. They find that for stocks in the largest quintile, a one-unit increase in AT decreases quoted spreads by 0.53 basis points. Several papers use data from the Deutsche Bourse, where the source of the trade (human or computer) is known. Using these data, Hendershott and Riordan (2009) find that algorithmic trading supplies liquidity when it is costly and demands liquidity when it is cheap. They find that AT also places efficient quotes and tend to move the price towards the efficient price. Brogaard (2010) looks at the NASDAQ market and examines the effect of AT on various outcomes such as market quality and volatility. [32]

The rest of the paper is organized as follows: Section II lays out the model; Section III introduces the datasets and variables used in this paper; Section IV reports results from the regressions and Section V concludes.

3.2 Model

The basic model of liquidity and AT is written as follows:

$$Spread_{it} = \alpha + \beta\, AT_{it} + \boldsymbol{\delta}\mathbf{X}_{\boldsymbol{t}} + \epsilon_{it}$$

where $ESPR_{it}$ is the relative effective spread, a measure of stock illiquidity, AT_{it} is a measure of algorithmic trading, and $\mathbf{X}_{\boldsymbol{t}}$ is a vector of control variables including innovations in aggregate market liquidity (as developed in Pastor and Stambaugh (2003)), the CBOE VIX index (a measure of market volatility), and the log of size (defined as share price times number

[32] Chaboud et al. (2009) concentrate AT in foreign exchange markets. There are also various reports on the Flash Crash (See Kirilenko et al. (2011) and SEC-CFTC (2010) report on the crash) that shed light on AT during an extreme event.

of shares outstanding). Since AT is not generally observed, we use several proxy measures. The first is the average time to fill an order on the NYSE. The second is the proportion of orders filled within fewer than 10 seconds from order placement. These measures can proxy for AT because an increase in the presence of algorithmic traders should lead to faster order fill times and subsequently increase the number of orders filled within ten seconds of order arrival. There are two potential reasons for why this may be true: (i) automated algorithms have the advantage of making quicker buy or sell decisions than human traders, and (ii) if even a small portion of the new traders using AT are high-frequency traders who use algorithms that monitor the market far more frequently, then in a given time period we can expect the number of participants in a stock to increase (human or otherwise) with the use of AT. While it is not possible to test these scenarios with the current data, these are testable implications that can potentially add to the discussion on the effect of AT in the marketplace.

The coefficient on AT, β, gives the effect of an increase in the incidence of such trading on stock liquidity. If β is positive, then AT tends to destabilize markets by increasing spreads; and if it is negative, then AT leads to increased market efficiency by narrowing spreads. However, ordinary least squares estimates of β in Equation 1 are unlikely to be consistent since the degree of AT is likely to be a function of the spread itself. For example, a trader may eschew AT due to the risk involved in markets turning against her position during times of uncertainty when markets are volatile and spreads are high. In this case, AT will be correlated with the error term. To circumvent this problem, we follow Hendershott et al. (2011) and instrument for AT using the NYSE's adoption of the auto quote system. The first stage equation is as follows:

$$AT_{it} = \gamma + \pi\, Q_{it} + \boldsymbol{\psi}\mathbf{X}_{\boldsymbol{t}} + \nu_{it}$$

where Q_{it} is an indicator variable for whether quotes were automatically disseminated for stock *i* on day *t*. A detailed discussion of this variable and proxies for AT can be found in the following section.

3.3 Data

The full sample consists of 661 common stocks (share code 10 or 11) listed on the NYSE between June 2001 and July 2005. Data used for this paper is from SEC-mandated Disclosure of Order Execution (DOE) and from NYSE's Trade and Quotes (TAQ) dataset. The DOE dataset contains information on order execution and routing practices across all U.S. market centers. These centers make monthly, electronic disclosures of basic information concerning quality of executions on a stock-by-stock basis such as the average price improvement on executed orders, time taken to execute orders of various sizes, etc. The following subsections document the construction of the liquidity measure, the autoquote variable, and the two measures of AT. To control for market liquidity and market volatility, we use the monthly liquidity measure proposed by Pastor and Stambaugh (2003) and CBOE VIX index, respectively.

3.3.1 Effective Spread

Intra-day bid and offer quotes are obtained in the same way as that of the stock price. The effective spread is defined as twice the absolute difference between the actual trade price and the midpoint of the bid-offer spread. Since quote updates from a trade are frequently recorded ahead of the trade, we cannot match the current trade price with the current quote. The trade price is matched with the quote in effect 5 seconds before the trade time-stamp as advocated in Lee and Ready (1991). Using these quotes, the relative effective spread for a given stock *i* at time *s* on day *t* is:

$$Relative\ Effective\ Spread_{its} = \frac{2 \times \left| P_{its} - \left(\frac{Ask_{its} + Bid_{its}}{2}\right) \right|}{\left(\frac{Ask_{its} + Bid_{its}}{2}\right)}$$

Effective spreads are a better proxy for liquidity than quoted spreads (the difference between bid and offer prices) for several reasons. Lee and Ready (1991) show that 30% of trades occur inside the spread and quoted spreads cannot capture the effect of such trades. Moreover, quotes are often "indications" to trade and may not be firm offers of a trade. Thus, effective spreads better captures the cost of a round-trip order by accounting for any price movements caused by market makers' arrival to fill the order at a better price. To calculate the spread as a percentage of the underlying stock price, we divide by the bid-offer midpoint or the price.[33] Anomalous observations that satisfied the following conditions are removed from the sample[34]:

1. Quoted Spread = Ask – Bid < 0;
2. Effective Spread / Quoted Spread > 4; (4 or 0.4? You have 0.4 below.)

Relative effective spreads greater than 0.4 are set to equal 0.4 so that results are not influenced by anomalous spreads. The daily effective spread is obtained by averaging the spread over times *s* within day *t*. The empirical analysis uses proportional effective spread which is calculated as the ratio of daily effective spread and midpoint of the bid-ask spread:

$$Spread_{it} = \frac{1}{n_{it}} \sum_{s=1}^{n_{it}} Relative\ Effective\ Spread_{its},$$

where n_{it} is the number of observations available on day *t* for stock *i*.

[33] Dividing either by price or the bid-offer midpoint did not affect the results. Expressing spread as a percentage can be useful to compare spreads across stock with different trading prices.
[34] Such filters are used in the empirical literature (see, for example.,Chordia, Sarkar, and Subrahmanyam (2005)) to remove the effects of incorrect data entries.

3.3.2 Proxies for AT

For a stock i in month m the average order fill time is calculated as follows[35]:

$$Fill\ Time_{im} =$$

$$\frac{5 * NSX_{im}^{0-9} + 20 * NSX_{im}^{10-29.9} + 45 * NSX_{im}^{30-59.9} + 180 * NSX_{im}^{60-299.9} + 1050 * NSX_{im}^{5-30m}}{Total\ No.of\ Shares\ Executed\ in\ month\ m}$$

where NSX^{0-9} is the number of shares executed in less than 10 seconds, $NSX^{10-29.9}$ is the number of shares executed between 10 and 29.9 seconds, $NSX_{im}^{30-59.9}$ is the number of shares executed between 30 and 59.9 seconds, $NSX_{im}^{60-299.9}$ is the number of shares executed between 60 and 299.9 seconds, and NSX_{im}^{5-30m} is the number of shares executed between 5 and 30 minutes . The second measure, the proportion of shares executed in less than 10 seconds is calculated as:

$$Prop_{im}^{0-9} = \frac{NSX_{im}^{0-9}}{Total\ Number\ of\ Shares\ Executed\ in\ month\ m}$$

3.3.3 Autoquote Variable:

Automated quotation dissemination was introduced in a staggered manner for NYSE stocks. The program began in January 2003 and ended on 27 May 2003, at which point quotes were being automatically disseminated for all NYSE stocks. The autoquote list is obtained from Prof. Hendershott's website. Section III of Hendershott et all (2011) describes the entire program in detail and discusses the importance of the staggered nature of the autoquote program that helps identify the effect of AT on liquidity.

---

[35] Data is available by order type (marketable and non-marketable) and by order size (number of shares per order.) The fill time is calculated for each order type and size first. Then, we average over all the various combinations of types and sizes to get a single measure of fill time for one stock in one month.

3.3.4 Summary Statistics

Tables 1 and 2 display the summary statistics for the main variables used in this paper. Table 2 reports the summary statistics for the whole sample. The average time taken to fill an order over this sample is 125 seconds or about 2 ½ minutes. The mean value for the proportion of orders filled in less than 10 seconds is 28%. The average stock is a medium-cap stock with a market value of $7 billion. For the whole sample, the average effective spread normalized by stock price is 0.6%. Table 2 reports the summary statistics for each market cap. On average, smaller stocks tend to have higher time to fill an order and the proportion of order filled within ten seconds increases as stock size increases. As expected, effective spreads also decrease by market –cap. . So, stock size appears to be related to both liquidity and AT, which is why we control for it in all our regressions.

3.4 Regression Model and Results

3.4.1 Regressions with average fill time as the AT proxy:

We first run the following simple fixed effects model:

$$ESPR_{it} = \alpha_i + \beta\ Fill\ Time_{it} + \boldsymbol{\delta}\mathbf{X_t} + \epsilon_{it}$$

Tables 3 and 4 display the results from the regressions above with and without the control variables for the full sample and for the samples split by market capitalization. The full sample is split by market cap as “micro-,” “small-,” “mid-,” and “large-” cap.[36] In any liquidity study, inclusion of micro- and small-cap stocks can be problematic as they are highly illiquid by nature and can drive the results when we pool across market caps. Thus, it is imperative to run

[36] A stock’s average market value (price times the number of shares outstanding) was used to split the sample by market capitalization. Micro cap stocks are those with an average market value of $250 million or less, small cap stocks lie within $250 million and $2 billion (not inclusive), mid-cap stocks are between $2 billion and $10 billion, and large-cap stocks are those with average market value of $10 billion and above.

the main regressions separately on each market-cap sample to see which stocks drive the results. Average time taken to fill orders increase with wider spreads, as evidenced by the positive slope in the small and mid-cap, and the full sample. The magnitude of the effect is quite small (0.0002%) and we are not sure which way the causation runs, since we would expect that any measure of algorithmic trading would be correlated with the error term. Including the control variables confirms the positive correlation, though only in small cap and full samples. Now, we hypothesize that the introduction of auto quote helps algorithms pick up market signals more quickly. This can lead to quicker order fulfillment in the market, thus decreasing average time to fill an order. So, in a first-stage regression of average fill times on auto quote, the coefficient on auto quote should be negative. In Table 5, the coefficient on autoquote is indeed negative and significant at the 1% level. A coefficient of -4.0585 can be interpreted as the decrease in the number of seconds an average order takes to get filled on the NYSE with the introduction of autoquotes. In the first stage, we see that autoquote had a significant effect on average time taken to fill an order for all but micro-cap stocks. The magnitude of the effect is higher for small and large cap stocks, with coefficients of -1.595 and -2.016, respectively, than mid-cap stocks (-0.697). Controlling for size, volatility, and market liquidity does not change the results, as seen in Table 7. The coefficient on size is negative, signifying that larger stocks tend to have deeper markets that aid in quicker order fill times than smaller stocks. This relationship is robust even when the regressions are run on samples split by market cap. For example, within small-cap stocks, a 1% increase in the natural logarithm of size decreases average time taken to fill an order by 0.15 seconds.

In the second stage, we can now estimate the causal effect of the AT proxy on effective spreads. In Table 6, the coefficient on average time to fill, a proxy for AT, is 0.0002; i.e. a one second decrease in time to fill an order is associated with a 2 bps (0.02%) decrease in spread. A quick examination of the split samples shows that this result is mostly driven by small-cap stocks. However, when we control for size, volatility, and market liquidity in Table 8, the fill time variable is significant only for mid-cap and full sample.

Thus far, we have used a measure of fill time which varies at the monthly level, whereas all other variables vary at the daily level. An alternative approach would be to average the daily variables up to the monthly level. So, all daily variables are averaged at the monthly level. Tables 9 through 11 display results from 2SLS regressions with the monthly variables.[37] Table 9 shows the results from the fixed effects regression of spread on fill time and the coefficient on fill time is not significant. Table 10 shows the results from the first stage of the 2SLS. Just like before, there is a negative relationship between the introduction of auto quotes and fill times. The introduction of autoquotes decreased fill times by 5.15 seconds per month. In the second stage results reported in Table 11, we see that there is no change in either magnitude or significance from before, and a one second decrease in the average fill time decreases effective spreads by 0.02% per month.

3.4.2 Regressions with $\boldsymbol{Prop_{it}^{0-9}}$ as the AT proxy:

Regressions were repeated with the second proxy for AT, $Prop_{it}^{0-9}$, the proportion of order filled between zero and nine seconds of order receipt. This measure is constructed as the ratio of the number of orders executed/filled between 0 and 9 seconds and the total numbers of orders executed. We hypothesize that orders in the market are filled at a faster rate with an increase in the use of AT. Thus, in the first stage regression, we would expect a positive coefficient on the autoquote variable. Tables 14 and 16 show the results from the first stage regression across the full and split samples. In the full sample, we do see a positive and significant coefficient on autoquote, though this result is not statistically significant across the

[37] While regressions were run on both split samples and the full sample, only results from the full sample regression are reported here because the coefficients of interest in both first and second stage were not significant even at 10% level.

split samples. As stock size increases, we see that more orders are filled in fewer than 10 seconds. This is perhaps due to the fact that the market for larger stocks tends to be liquid with increasing probabilities that we can always find participants willing to take opposing positions for a given order. In the second stage results reported in Table 15, a 1% increase in the proportion of orders filled in fewer than 10 seconds decreases spreads by 0.62%. In Table 17, the control variables take away the statistical significance of the effect of AT, though the signs are the same. These effects are much stronger than in a simple fixed effects OLS models, such as those in Tables 12 and 13, where we do not instrument for AT. A 1% increase in AT in the fixed effects OLS model decreases spreads by only 0.0044%.

In general, the two measures proposed in this paper seem to be quite useful proxies for AT. We find that there is a significant difference in the magnitude between the coefficient on AT in an OLS setting and the coefficient on AT in a 2SLS model setting. Since the 2SLS coefficient is greater than the OLS coefficient, by definition, there is a negative correlation between unobservable determinants of stock illiquidity and algorithmic trading. This means that, when there exists high levels of idiosyncratic illiquidity in markets, we would expect to see fewer algorithmic traders as they want to avoid being picked-off. This relationship could be related to the results in Hendershott and Riordan (2011), who examine algorithmic trades in the Deutsche Boerse. Using a probit model, they find that algorithmic trades consume (demand) liquidity when spreads are narrow (wide.)

3.5 Conclusion

In this paper, we propose two measures of AT—the average time taken to fill orders on the NYSE and the proportion of orders executed within ten seconds of order arrival. These measures are constructed from a rarely used finance dataset that has information on order execution and quality on the various U.S. stock market centers. Using these measures as a proxy for the latent phenomenon of algorithmic trading, we find that increases in AT decreases spreads and improves stock liquidity. A 1 second decrease in average time taken to fill an order decreases spreads by 2 basis points (0.02 %).

Algorithmic and high –frequency trading practices have been in the news recently and there is a lot of debate about whether they are beneficial to efficient markets. The nature of this debate is also a political one since there are calls for taxes on financial transactions. The empirical literature on the effects of AT on market quality and liquidity is sparse, and this paper adds to that literature in providing two proxies for AT and shows that AT can be beneficial in markets. Under this scenario where AT and by way, high-frequency trading may be beneficial, a financial transactions tax may hinder market efficiency and capital formation. More attention is certainly needed in this area of research and since this dataset stops in 2005, it would be interesting to study the evolution of AT and its effect on market quality since then in equity markets. An event study around the financial crisis would also help shed some light on counterfactuals such as how markets would have fared without such traders.

Table 3.1: Summary Statistics (Full sample)

	Obs	Mean	Std. Dev.	Min	Max
$\boldsymbol{Fill\,Time_{it}}$	560,330	125.73	51.08	16.59189	273
$\boldsymbol{Prop_{it}^{0-9}}$	560,330	0.28	0.12	0	0.765884
$\boldsymbol{Size_{it}}$	560,330	705.03	19,100	13.19	204,000
$\boldsymbol{Spread_{it}}$	560,330	0.006	0.019	0	0.4
$\boldsymbol{Q_{it}}$	560,330	0.53	0.49	0	1

Table 3.2: Summary Statistics (By Market Cap)

	Obs	Mean	Std. Dev.	Min	Max
			MICRO-CAP		
$Fill\ Time_{it}$	55,632	163.36	50.51	17.06	273.00
$Prop_{it}^{0-9}$	55,632	0.20	0.09	0.00	0.77
$Size_{it}$	55,632	147.77	56.09	13.19	240.38
$Spread_{it}$	55,632	0.02	0.03	0.00	0.40
Q_{it}	55,632	0.47	0.50	0.00	1.00
			SMALL-CAP		
$Fill\ Time_{it}$	252,514	148.92	46.09	16.59	273.00
$Prop_{it}^{0-9}$	252,514	0.26	0.11	0.01	0.62
$Size_{it}$	252,514	954.60	493.94	250.75	1,979.34
$Spread_{it}$	252,514	0.01	0.02	0.00	0.40
Q_{it}	252,514	0.53	0.50	0.00	1.00
			MID-CAP		
$Fill\ Time_{it}$	161,729	105.88	36.14	36.54	273.00
$Prop_{it}^{0-9}$	161,729	0.32	0.11	0.03	0.66
$Size_{it}$	161,729	4,418.07	2,071.97	2,007.81	9,935.52
$Spread_{it}$	161,729	0.00	0.01	0.00	0.40
Q_{it}	161,729	0.54	0.50	0.00	1.00
			LARGE-CAP		
$Fill\ Time_{it}$	90,455	73.37	22.62	29.96	190.55
$Prop_{it}^{0-9}$	90,455	0.37	0.12	0.08	0.71
$Size_{it}$	90,455	33,000.00	37,800.00	10,400.00	204,000.00
$Spread_{it}$	90,455	0.00	0.02	0.00	0.40
Q_{it}	90,455	0.55	0.50	0.00	1.00

Note: A stock's average market value (price times number of shares outstanding) was used to split the sample by market capitalization. Micro cap stocks are those with an average market value of $250 million or less, small cap stocks lie within $250 million and $2 billion (not inclusive), mid-cap stocks are between $2 billion and $10 billion, and large-cap stocks are those with average market value of $10 billion and above.

Table 3.3: The Effect of Average Time to Fill on Spread (OLS)

	Market Cap			
	Small	Mid	Large	All
$Fill\ Time_{it}$	6.22e-06***	7.21e-06***	-1.04E-05***	2.51E-06***
	[1.10e-06]	[1.23e-06]	[3.83e-06]	[1.00e-06]
Number of Stocks	299	182	98	661
Number of Observations	255,164	155,533	86,319	560,330
R-squared	0.2	0.21	0.16	0.23

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.4: The Effect of Average Time to Fill on Spread (OLS with Controls)

	Market Cap			
	Small	Mid	Large	All
$Fill\ Time_{it}$	2.30e-06**	4.69e-07	-3.24e-06	-1.97e-06**
	[1.10e-06]	[1.22e-06]	[3.92e-06]	[1.00e-06]
VIX_t	.0001***	.0001***	.0001	0.0001***
	[.00001]	[.00001]	[.00003]	[0.00001]
$Size_{it}$	-.0046***	-.0033***	.0021	-0.0053***
	[.0001]	[.00012]	[.0002]	[0.0001]
L_t^{Mkt}	.0224***	-.0091**	-.0049	-0.0024
	[.0078]	[.0043]	[.0049]	[0.0048]
Number of Stocks	299	182	98	661
Number of Observations	255,164	155,533	86,319	560,330
R-squared	0.2	0.22	0.16	0.24

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.5: The Effect of Autoquote on Average Time to Fill

	Market Cap			
	Small	Mid	Large	All
Q_{it}	-1.5955***	-0.697*	-2.016***	-4.0585***
	[0.6118]	[0.4098]	[0.2342]	[0.3284]
Number of Stocks	299	182	98	661
Number of Observations	255,164	155,533	86,319	560,330
R-squared	0.49	0.70	0.80	0.65

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.6: The Effect of Average Time to Fill on Spread (2SLS)

	Market Cap			
	Small	Mid	Large	All
$Fill\ Time_{it}$	0.0006**	0.001	-0.00019	0.0002***
	[0.00029]	[0.00080]	[0.00015]	[.00004]
Number of Stocks	299	182	98	661
Number of Observations	255,164	155,533	86,319	560330

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.7: The Effect of Autoquote on Average Time to Fill (with Controls)

	Market Cap			
	Small	Mid	Large	All
Q_{it}	-0.8567	-1.0698***	-1.8419***	-3.6288***
	[.6090]	[.4063]	[.2325]	[.3257]
VIX_t	-0.1073***	-0.0784***	-0.0406***	-0.0744***
	[.0321]	[.0232]	[.0154]	[.0198]
$Size_{it}$	-14.6008***	-9.6072***	-5.4884***	-10.5218***
	[.2228]	[.1805]	[.1756]	[.1553]
L_t^{Mkt}	120.32***	74.38***	-62.89***	83.5499***
	[14.9911]	[13.70]	[8.32]	[9.2746]
Number of Stocks	299	182	98	661
Number of Observations	255,164	155,533	86,319	560,330
R-squared	14.99	0.70	0.80	0.66

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.8: The Effect of Average Time to Fill on Spread (2SLS with Controls)

	Market Cap			
	Small	Mid	Large	All
$Fill\ Time_{it}$	0.00096	0.0007*	-0.00018	0.0002***
	[0.0007]	[0.00042]	[0.00017]	[0.00005]
VIX_t	0.0002**	0.00018***	0.0014*	0.00013***
	[0.00008]	[0.00004]	[0.0009]	[0.000014]
$Size_{it}$	0.0094	0.0047	0.0014	-0.003***
	[0.0105]	[0.0041]	[0.0009]	[0.0005]
L_t^{Mkt}	-0.0941	-0.0688**	-0.0154	-0.0205***
	[0.0884]	[0.0339]	[0.0116]	[0.0069]
Number of Stocks	299	182	98	661
Number of Observations	255,164	155,533	86,319	560,330

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.9: The Effect of Average Fill Time on Spread (OLS and Monthly Regressions)

	[1]	[2]
$Fill\ Time_{im}$	4.26E-06	-4.7E-07
	[3.01e-06]	[2.86e-06]
VIX_m		0.0011
		[0.0007]
$Size_{im}$		-0.0056***
		[0.0003]
L_m^{Mkt}		-0.0221
		[0.0199]
Number of Stocks	661	661
Number of Observations	27208	27208
R-squared	0.4899	0.5017

Table 3.10: The Effect of Autoquote on Average Time to Fill (Monthly Regressions)

Variable	[1]	[2]
Q_{it}	-5.6512***	-5.1489***
	[1.6468]	[1.6330]
VIX_t		2.9329*
		[1.5151]
$Size_{it}$		-10.47***
		[0.7354]
L_t^{Mkt}		31.24
		[51.98]
Number of Stocks	661	661
Number of Observations	27,208	27,208
R-squared	0.65	0.66

Table 3.11: The Effect of Average Time to Fill on Spread (2SLS and Monthly Regressions)

Variable	[1]	[2]
$Fill\ Time_{im}$	0.0002**	0.0002*
	[0.0001]	[0.0001]
VIX_m		0.0004
		[0.0009]
$Size_{im}$		-0.0031**
		[0.0014]
L_m^{Mkt}		-0.0296
		[0.0237]
Number of Stocks	661	661
Number of Observations	27,208	27,208

Table 3.12: The Effect of $Prop_{it}^{0-9}$ on Spread (OLS)

	Market Cap			
Variable	Small	Mid	Large	All
$Prop_{it}^{0-9}$	-0.0055***	-0.0015***	-0.0020***	-0.0044***
	[0.0006]	[0.0005]	[0.0007]	[0.0004]
Number of Stocks	299	182	98	661
Number of Observations	252,514	161,729	90,455	560,330
R-squared	0.20	0.22	0.17	0.23

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.13: The Effect of $Prop_{it}^{0-9}$ on Spread (OLS with Controls)

	Market Cap			
Variable	Small	Mid	Large	All
$Prop_{im}^{0-9}$	-0.0035***	-0.0016***	-0.0021***	-0.0034***
	[0.0006]	[0.0005]	[0.0007]	[0.0004]
VIX_t	0.0001***	0.0001***	0.0001***	0.0001***
	[0.00001]	[0.00001]	[0.00003]	[0.0001]
$Size_{it}$	-0.0046***	-0.0033***	0.0021***	-0.0052***
	[0.0001]	[0.0001]	[0.0002]	[0.0001]
L_t^{Mkt}	0.0228***	-0.0089***	-0.0048	-0.0026
	[0.0078]	[0.0043]	[0.0049]	[0.0048]
Number of Stocks	299	182	98	661
Number of Observations	252,514	161,729	90,455	560,330
R-squared	0.20	0.22	0.17	0.24

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.14: The Effect of Autoquote on $Prop_{it}^{0-9}$

	Market Cap			
Variable	Small	Mid	Large	All
Q_{it}	0.0023**	-0.0065***	0.0003	0.0016***
	[0.0011]	[0.0014]	[0.0015]	[0.0007]
Number of Stocks	299	182	98	661
Number of Observations	252,514	161,729	90,455	560,330
R-squared	0.69	0.76	0.79	0.75

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.15: The Effect of $Prop_{it}^{0-9}$ on Spread (2SLS)

	Market Cap			
Variable	Small	Mid	Large	All
$Prop_{im}^{0-9}$	-0.5093*	0.0386	-1.25	-0.6235**
	[0.2625]	[0.0542]	[5.99]	[0.2911]
Number of Stocks	299	182	98	661
Number of Observations	252,514	161,729	90,455	560,330

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.16: The Effect of Autoquote on $Prop_{it}^{0-9}$ (with Controls)

	Market Cap			
Variable	Small	Mid	Large	All
Q_{it}	0.0008	-0.0065***	0.0002	0.0012*
	[0.0011]	[0.0014]	[0.0015]	[0.0007]
VIX_t	0.0001***	-3.82e-06	0.00002	0.00007**
	[0.00005]	[0.00006]	[0.00008]	[0.00003]
$Size_{it}$	0.0202***	-0.0003	0.0033***	0.0095***
	[0.00045]	[0.0006]	[0.0008]	[0.0003]
L_t^{Mkt}	0.0413	0.0557	-0.0907	-0.0046
	[0.0320]	[0.03849]	[0.0460]	[0.0217]
Number of Stocks	299	182	98	661
Number of Observations	252,514	161,729	90,455	560,330
R-squared	0.69	0.76	0.79	0.75

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

Table 3.17: The Effect of $Prop_{it}^{0-9}$ on Spread (2SLS with Controls)

Variable	Market Cap Small	Mid	Large	All
$Prop_{im}^{0-9}$	-0.9181	0.0603	-1.8096	-0.6158
	[1.19]	[0.0551]	[11.88]	[0.3763]
VIX_t	0.0002	0.0001***	0.00024	0.0001***
	[0.0001]	[0.00001]	[0.0003]	[0.00003]
$Size_{it}$	0.0138	-0.0033***	0.0082	0.0005
	[0.0241]	[0.0001]	[0.0399]	[0.0035]
L_t^{Mkt}	0.0606	-0.0124**	-0.1688	-0.0054
	[0.0579]	[0.0057]	[1.08]	[0.0143]
Number of Stocks	299	182	98	661
Number of Observations	252,514	161,729	90,455	560,330

Note: Each column represents a separate regression. Each regression includes fixed effects for stock, year, month, year*month, and day of the week.

3.6 References

Boehmer, Ekkehart, 2005, "Dimensions of execution quality: Recent evidence for US equity markets," *Journal of Financial Economics* 78(3), pp. 553-582.

Bessembinder, Hendrik, 2003, "Issues in assessing trade execution costs," *Journal of Financial Markets* 6(3), pp. 233–257.

Brogaard, Jonathan, 2010, "High Frequency Trading and Market Quality," Working Paper.

Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2009, "Rise of the machines: Algorithmic trading in the foreign exchange market," Working paper, Federal Reserve Board.

Hasbrouck, Joel, and Gideon Saar, 2011, "Low-Latency Trading," Working Paper.

He, Chen, Elizabeth Odders-White, and Mark J. Ready, 2006, "The impact of preferencing on execution quality," *Journal of Financial Markets* 9(3), pp. 246–273.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, "Does Algorithmic Trading Improve Liquidity?," *Journal of Finance* 66(1), pp. 1–33.

Hendershott, Terrence, and Ryan Riordan, 2011, "Algorithmic Trading and Information," NET Institute Working Paper No. 09-08.

Kirilenko, Andrei A., Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, "The Impact of High Frequency Trading on an Electronic Market," Working Paper, May 26, 2011.

Pastor, Lubos, and Robert F. Stambaugh, 2003, "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy* 111(3), pp. 642-685.

U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commissions, September 30, 2010, "Finding Regarding the Market Events of May 6, 2010."